

nature



THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

ATOMTRONICS REVISITED

Electronics-style hybridology in an ultracold quantum gas **PHYSICS** **444** **101**

**PROBABILISTIC
PROBLEMS**

Why it's so hard to
predict the future

448-450

**WHAT'S THE
MATTER?**

How the brain
decides what's
real

460-462

**GENE DE
AMERICA**

How the genome
shapes the nation

470-472

CLIMATE CHANGE

How the world
is changing

480-482

THIS WEEK

EDITORIALS

ECOLOGY The troubled tale of the wolf, the moose and the ice **p.132**

WORLD VIEW German court verdict ends scientist's surreal nightmare **p.133**

PESTICIDES Evidence of chemical damage to bumblebees **p.134**



Suicide watch

Despite a high death toll, public-health efforts to combat suicide lag far behind those focused on preventing accidents and diseases such as cancer. A US initiative aims to redress the balance.

Some 38,000 people killed themselves in the United States in 2010. That's more than were killed in traffic accidents (34,000) or by prostate cancer (29,000), and more than twice the number murdered (16,000). Shocking though that is, many other countries monitored by the Organisation for Economic Co-operation and Development have even higher suicide rates. So why do public-health authorities put less effort into preventing death from suicide than they do death from accidents or diseases such as prostate cancer?

One institution that has started to take the matter very seriously is the US army. Since 2008, the suicide rate among soldiers has exceeded that of the general population, and in the past few years the army has lost more soldiers to suicide than to combat. In 2009, the army launched a US\$65-million, six-year project called Army STARRS to collect genomic, medical, psychological and lifestyle data from more than 100,000 soldiers to try to identify suicide risk factors and prevention measures, as well as biomarkers of resilience such as epigenetics or brain connectivity. In 2010, it co-launched the National Action Alliance for Suicide Prevention, a public-private partnership, which last week released a pioneering 172-page report on suicide and how it might be tackled.

The report, produced and published in partnership with the US National Institute of Mental Health in Bethesda, Maryland, outlines a strategy to reduce suicide rates in the general population by 20% over the next five years. It also makes shockingly clear how little is known about suicide. There is no standard way to define and so recognize what it means to be suicidal. Relevant statistics are not routinely collected, which makes it hard to know, for example, the effect of round-the-clock crisis teams, and good follow-up care for those who attempt suicide.

Cases of suicide linked to cyber-bullying in young people feature prominently in the media, but few studies have addressed how social media might increase suicide risk through bullying or contagion (prompting copycat suicides). In any case, people over the age of 65 kill themselves much more frequently than do young people.

Two things we do know. First, a high number of people with a psychiatric disorder such as schizophrenia, depression or substance abuse kill themselves — somewhere between 50% and 90% of all suicides are thought to be associated with mental illness. Second, stressful life events, particularly during childhood, greatly increase suicide risk. However, most people who are under stress or mentally ill do not kill themselves. And even as the use of psychiatric medications has soared in the past two decades, suicide rates in the United States and most other countries have remained stable. So what is going on? And what might help?

It will never be possible to eliminate suicide, but it should be possible to reduce rates in different risk groups by attacking the problem from many sides. Biological approaches could identify and help the vulnerable, and sociological interventions could reduce stress in societies.

More long-term studies such as Army STARRS are required to shed light on the biology. And clinical trials can identify therapies that target personality traits or feelings likely to lead to suicide — impulsivity and

helplessness, perhaps. One large clinical trial that directly addressed suicide and psychiatric disease indicated that the antipsychotic drug clozapine could help to cut suicide rates in people with schizophrenia (L. Alphas *et al. Schizophr. Bull.* **30**, 577–586; 2004). And small trials have hinted that lithium may do the same for those with depression.

There are no good animal models for suicide risk at present, so biological investigations will have to rely on work with humans. But much can already be done to reduce suicide numbers, even in the absence of biomarkers. One powerful option, on which the report's strategy for reducing suicides by 20% strongly depends, would be to

reduce people's access to means of suicide.

Surprisingly, many people intent on suicide abandon their plan if their chosen means is not available. Firearms account for about half of US suicide deaths, and modelling work carried out for the new report shows that almost 10% of all suicides could be prevented by

restricting access to guns. In 2010, 735 people

in the United States killed themselves with carbon monoxide from car exhausts; the report suggests that 600 of those deaths might have been prevented if car manufacturers were required to install a sensor inside the vehicle that turns off the engine when carbon monoxide builds up.

The report's 20% target will probably not be achieved in the desired five years, but it opens a useful debate that will help more people to understand that the action of committing suicide needs to be considered in the same way as a disorder — as something that can be addressed, not an unavoidable product of circumstance. ■

"There is no standard way to define and so recognize what it means to be suicidal."

Number crunch

The correct use of statistics is not just good for science — it is essential.

In the fragmented media marketplace, it is a brave publisher that dismisses the professional competence of most of its readers. So sensitive subscribers might want to avoid page 150 of this week's *Nature*.

The criticism in question appears in a News Feature on the thorny issue of statistics. When it comes to statistical analysis of experimental data, the piece says, most scientists would look at a *P* value of 0.01 and "say that there was just a 1% chance" of the result being a false alarm. "But they would be wrong." In other words, most researchers do not understand the basis for a term many use every day. Worse, scientists misuse it. In doing so, they help to bury scientific truth beneath an avalanche of false findings that fail to survive replication.

As the News Feature explains, rather than being convenient shorthand for significance, the P value is a specific measure developed to test whether results touted as evidence for an effect are likely to be observed if the effect is not real. It says nothing about the likelihood of the effect in the first place. You knew that already, right? Of course: just as the roads are filled with bad drivers, yet no-one will admit to driving badly themselves, so bad statistics are a well-known problem in science, but one that usually undermines someone else's findings.

The first step towards solving a problem is to acknowledge it. In this spirit, *Nature* urges all scientists to read the News Feature and its summary of the problems of the P value, if only to refresh their memories.

The second step is more difficult, because it involves finding a solution. Too many researchers have an incomplete or outdated sense of what is necessary in statistics; this is a broader problem than misuse of the P value. Among the most common fundamental mistakes in research papers submitted to *Nature*, for instance, is the failure to understand the statistical difference between technical replications and independent experiments.

Statistics can be a difficult discipline to master, particularly because there has been a historical failure to properly teach the design of experiments and the statistics that are relevant to basic research. Attitudes are also part of the problem. Too often, statistics is seen as a service to call on where necessary — and usually too late — when, in fact, statisticians should be involved in the early stages of experiment design, as well as in teaching. Department heads, lab chiefs and senior scientists need to upgrade a good working knowledge of statistics from the 'desirable' column in job specifications to 'essential'. But that, in turn, requires universities and funders to recognize the importance of statistics and provide for it. *Nature* is trying to do its bit and to acknowledge its own shortcomings. Better use of statistics is a central

plank of a reproducibility initiative that aims to boost the reliability of the research that we publish (see *Nature* **496**, 398; 2013). We are actively recruiting statisticians to help to evaluate some papers in parallel with standard peer review — and can always do with more help. (It has been hard to find people with the right expertise, so do please get in touch.) Our sister journal *Nature Methods* has published a series of well-received columns, Points of Significance, on statistics and how to use them.

"Too many researchers have an incomplete or outdated sense of what is necessary."

Some researchers already do better than others. In the big-data era, statistics has changed from a way to assess science to a way of doing science — and some fields have embraced this. From genomics to astronomy, important discoveries emerge from a mass of information only when they are viewed through the correct statistical prism. Collaboration between astronomers and statisticians has spawned the discipline of astrostatistics. (This union is particularly apposite, because it mirrors the nineteenth-century development of statistical techniques such as least squares regression to solve problems in celestial mechanics.)

Among themselves, statisticians sometimes view their contribution to research in terms of a paraphrase of chemical giant BASF's classic advertising tag line: "We don't make the products. We make them better." In doing so, they sell themselves short. Good statistics can no longer be seen as something that makes science better — it is a fundamental requirement, and one that can only grow in importance as funding cuts bite and competition for resources intensifies.

Most scientists use statistics. Most scientists think they do it pretty well. Are most scientists mistaken about that? In the News Feature, *Nature* says so. Go on, prove us wrong. ■

Lone wolves

A declining island wolf population underlines the influence that humans have on nature.

Ecologists have studied the wolves and moose on Isle Royale, a remote island in Lake Superior, for more than 50 years. As we report on page 140, after decades of isolation and inbreeding, the wolf population may be on the verge of dying out.

The US National Park Service, which manages the island, is moving slowly in deciding how to proceed. It has three options: total non-intervention; reintroduction of wolves only after the current population has hit zero; or pre-emptive genetic rescue by bringing in wolves from the mainland to diversify the gene pool. Arguments for non-intervention tend to rely on the perceived need to let nature take its course. This is nonsense. The whole system is highly artificial: wolves and moose have been on the island for less than 100 years, and human activity has been key to the wolves' decline. A previous wolf-population crash in the 1980s was caused by a disease transmitted by a domestic dog. Anthropogenic climate change is almost certainly reducing how often ice bridges form to the mainland, which makes it hard for new wolves to come to the island. Some even think that humans put moose on Isle Royale in the first place.

Arguments are more convincing for reintroducing wolves only if the current population dies out: waiting and watching may yield some useful insights into how highly inbred populations function. But the ecologists who run the island's predator-prey observation study warn that, as the wolves die out, the moose will gorge unchecked on their key food plant, balsam fir, preventing the plant from regenerating. The researchers think that by the time the old wolf population has died out and a new one is established, the ecosystem may have become

dominated by pine or spruce, without enough firs to support a moose population that can in turn feed a viable wolf population. If the wolves die out, they could become nearly impossible to reintroduce.

And that might be fine, except that tourists and locals love the wolves of Isle Royale, and the National Park Service was founded with an obligation to protect "the enjoyment of future generations". Furthermore, the predator-prey study — the world's longest — would have to end. That would be a shame: it would be difficult to find another place where none of the predators, herbivores or trees are routinely exploited by humans.

The study's lead ecologists are in favour of genetic rescue. This fairly cheap intervention would allow the project to continue, and would stabilize an ecosystem with which many people feel a deep connection. Some researchers have suggested that any data on reintroduced wolves would have to be treated with caution. Certainly, the influence of the reintroduction would be acknowledged and studied. But the introduced population would not be any more artificial than the population that survived disease, or that which could suffer the effects of climate change.

Isle Royale data help ecology to approach one of its grandest questions. As study leader John Vucetich puts it: "Are ecosystems like other physical systems, governed by law-like patterns and processes, or are they more like human history, where we see one contingency after the next?" The early years of the study seemed to support predictions that in a closed system, predator and prey populations would follow law-like mirror-image cycles, driven by predation pressure. But the data never fitted the theoretical curves that well. And since then, factors from disease to fir abundance, weather, moose ticks and wolf inbreeding have taken turns as the key driver in shaping the populations.

The driver that will shape the future of Isle Royale is now the decision on whether to stage a rescue. Thus of the story of all Earth's systems is writ small on a wooded isle in a frozen lake: the course of human history is no longer merely analogous to the course of ecology. Ecology depends on human history. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

STRUCTURAL BIOLOGY

Enzyme shifts shape to edit DNA

Structural biologists show how the DNA-cutting enzyme at the heart of a powerful gene-editing system changes shape to slice through DNA.

Martin Jinek at the University of Zurich, Switzerland, and his colleagues studied the CRISPR-Cas system, in which an RNA molecule is used to guide and program the Cas enzyme to break through a specific segment of DNA.

The team used X-ray crystallography and single-particle electron microscopy to explore the structure of Cas9 enzymes from two different bacterial species. They found that when the Cas9 enzymes bind to the guide RNA, the active sites rearrange to form a central channel that can attach to DNA.

Furthermore, the Cas9 enzyme is easily programmed by certain guide RNAs because of the amino-acid sequence at the enzyme's active site, which allows the enzyme's shape to be customized.

Science <http://doi.org/rcs> (2014)

ECOLOGY

Pesticides drag bumblebees down

Bumblebees exposed to pesticides bring pollen back to their colonies less often and have lower body mass than unexposed counterparts.
Hannah



Feltham at the University of Stirling, UK, and her colleagues fed bumblebees (*Bombus terrestris*; pictured) in the lab for two weeks on pollen and sugar water laced with doses of the pesticide imidacloprid that they might come across in the field. The researchers then attached tiny radio tags to the bees and placed colonies out in the field to monitor their foraging behaviour. They found that exposed insects carried pollen back to the colonies on 40% of trips, compared with more than 60% for unexposed bees.

Gemma Baron and her colleagues at Royal Holloway University of London in Egham, UK, tested the effect

of the pesticide λ -cyhalothrin. They discovered that pesticide-treated colonies produced smaller workers than unexposed colonies.

The researchers suggest that pesticides could contribute to a reduction in queen production and to bumblebee colony failure.

Ecotoxicology <http://doi.org/rct> (2014); *J. Appl. Ecol.* <http://doi.org/rcv> (2014)

MICROBIOLOGY

Glowing probe reveals infection

A fluorescent probe could help physicians to pinpoint where the dangerous bacterium

Staphylococcus aureus is living in a patient's body.

Some bacterial infections are difficult to locate directly, so James McNamara of the University of Iowa in Iowa City and colleagues created a molecular probe that reacts with an enzyme made by *S. aureus* called micrococcal nuclease. The enzyme cuts up small strings of nucleic acid molecules, and the probe consists of two nucleic acid bases attached to a fluorescent molecule.

The researchers injected the probe into a mouse with a *S. aureus* infection in its muscles. The bacterial enzyme cut the probe in two, activating the fluorescent molecule and



ZOOLOGY

Fish have it easy in schools

Fish might expend less energy swimming in a school than alone, probably by benefiting from wakes created by nearby fish.

Charlotte Hemelrijk at the University of Groningen in the Netherlands and her colleagues modelled the fluid dynamics of two-dimensional schools of animals, using various configurations of fish-like shapes. They found that fish in almost all the arrangements convert

power more efficiently into forward speed than do lone swimmers.

In some configurations, this occurs as fish swim behind one another and move their heads side to side. This means that they benefit from wakes from the preceding fish that speed them up, and are not impeded by swirling wakes that would slow them down.

Fish Fish. <http://doi.org/rb9> (2014)

BERNARD RADVANEY/CORBIS

REDMOND DURRELL/ALAMY

causing the bacteria to glow under a microscope.

Several other pathogens produce these nuclease enzymes, suggesting that the approach could be customized to detect specific species of bacteria.

Nature Med. <http://doi.org/rcf> (2014)

NEUROSCIENCE

How marijuana boosts eating

A specific receptor in the brain explains how hunger improves sense of smell and drives feeding behaviour in mice.

To probe the mechanism that links hunger, olfaction and eating, Giovanni Marsicano at the French National Institute of Health and Medical Research in Bordeaux and his colleagues focused on the CB₁ receptor, which is activated by a number of molecules, including THC, the active ingredient in marijuana. They found that the receptor is expressed in neurons that reach into the brain's olfaction centre. Blocking this receptor in mice that had fasted for 24 hours decreased their food intake. Stimulating the receptor with THC boosted sensitivity to odours and feeding in the hungry animals.

The CB₁ olfactory circuitry could offer a target for the treatment of eating disorders, the authors say.

Nature Neurosci. <http://dx.doi.org/10.1038/nn.3647> (2014)

MICROBIOLOGY

CRISPR takes out bacterial species

Researchers have used molecules involved in precise gene editing to selectively remove specific bacterial strains from mixed cultures, based on the microbes' DNA sequences.

The CRISPR gene-editing system uses RNA molecules from bacterial immune systems that recognize and cut foreign DNA. Chase Beisel and his colleagues at North Carolina State University in Raleigh engineered these molecules to target DNA sequences from individual strains of *Escherichia coli* and *Salmonella* in pure and mixed cultures. They showed that their strategy could distinguish between strains sharing more than 99% of their genetic content, and could remove specific bacteria.

This approach could one day lead to the development of antibiotics that pinpoint harmful bacteria and leave beneficial ones alone, the authors say.

mBio 5, e00928-13 (2014)

CRYOSPHERE

Glacier reaches record speed

Greenland's fastest-moving glacier has gained the highest flow speed ever observed in any sea-bound glacier in Greenland or Antarctica.

Jakobshavn Isbræ (pictured) in West Greenland has been retreating since the mid-1990s. Using high-resolution satellite radar observations to map the glacier's movements since 2009, Ian Joughin at the University of Washington in Seattle and his colleagues report that it has sped up by 30–50% in the past two years. In the summer of 2012, the glacier reached a record pace of more than 17 kilometres per year — four times greater than during the 1990s.

The Cryosphere 8, 209–214 (2014)

COMMUNITY CHOICE

The most viewed papers in science

MENTAL HEALTH

Greener spaces mean better living

HIGHLY READ
on Twitter in the past month

People who move from urban areas with less vegetation to places with more report improved mental health.

Mathew White and his colleagues at the University of Exeter Medical School in Truro, UK, used British Household Panel Survey data to analyse the change in mental health scores over five years for people who had moved from one urban residential area to another during that time period. The researchers controlled for changes in individual's lives, such as employment and marital status, and found that moving to a leafier area led to an immediate and sustained improvement in mental health.

People who moved to less green areas reported worse mental health, but returned to their baseline level after the move. *Environ. Sci. Technol.* 48, 1247–1255 (2014)

CLIMATE CHANGE

Pacific winds slow warming

An unprecedented strengthening of the Pacific trade winds over the past two decades has altered ocean circulation and could help to explain the current plateau in global temperatures.

Matthew England at the University of New South Wales in Australia and his colleagues used a global climate model to look at the impact of the stronger winds. Their model suggests that the fiercer trade winds help to cool the central and eastern Pacific while driving warm water into the deeper ocean in the west. They found that the wind effect cooled global temperatures by 0.1–0.2°C in 2012 compared to a control run without increased trade winds.

Models assessed by the Intergovernmental Panel on Climate Change have generally overestimated recent warming trends. However, the authors suggest that if stronger trade winds are factored in, the climate models tally with the observed temperature changes.

Nature Clim. Change <http://dx.doi.org/10.1038/nclimate2106> (2014)

MATERIALS

Asteroids prove hard to weld

Future space colonists hoping to make structures out of local materials will face a serious construction challenge: metal-rich asteroids are not suited to current welding techniques.

Electron-beam welding has been used in space to join materials such as steel, but it was unknown whether the technique would work on extraterrestrial materials. John Elmer and his colleagues at the Lawrence Livermore National Laboratory in California welded a fragment of the Canyon Diablo meteorite — a 57,000-tonne iron-rich rock that hit Arizona 50,000 years ago — using an electron beam under a vacuum. High levels of phosphorous in the meteorite created cracks as the weld cooled, and the high carbon content made the weld harder than the surrounding material, leading to a weak joint.

Sci. Technol. Weld. Join. <http://dx.doi.org/10.1179/1362171813Y.0000000188> (2014)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



KONRAD STEFFEN/CRES/UNIVERSITY OF COLORADO/AMAP

SEVEN DAYS

The news in brief

POLICY

UK animal research

The British government says that it is still committed to cutting the number of animals used in research, despite failing to meet a 2010 pledge to do so. The number of animal experiments in the United Kingdom topped 4 million in 2012, up from more than 2.5 million in 2000. On 7 February, science minister David Willetts unveiled a new plan to reduce animal research, which included encouraging data sharing and providing advice on alternatives to animal testing. But he stopped short of setting a numerical limit on animal experiments. See go.nature.com/zvmeru for more.

Suicide prevention

A national research agenda for suicide prevention in the United States was published on 5 February. The plan was drawn up by the National Action Alliance for Suicide Prevention, a public-private partnership. It sets a goal of decreasing suicides by 20% over five years. In its report, the team said that researchers should focus on the areas that

JOURNALISM GRANT

The International Development Research Centre in Ottawa is offering a science-journalism grant worth up to Can\$60,000 (US\$54,000). The winner will spend six months working in *Nature's* London newsroom and reporting science stories from developing countries. Candidates must be English-speaking Canadian citizens or permanent residents, and have at least three years' journalism experience. For more information, see go.nature.com/len5fl.



FREDERIC STEVENS/GETTY

Illegal ivory crushed

France destroyed 3 tonnes of poached ivory on 6 February. It is the first European country to take such a step since the global ban on ivory came into force in 1989. "We are resolved to

continue the fight against trafficking and to remove any temptation to recover the seized ivory" for black-market sales, said French environment minister Philippe Martin.

will prevent the most deaths, such as identifying at-risk people through mental-health screenings in hospital emergency departments, and preventing suicidal people from accessing firearms. See page 131 for more.

Wolf plan flawed

The US government's proposal to weaken protection for grey wolves (*Canis lupus*) is not based on good science, said an independent review panel on 7 February. The US Fish and Wildlife Service (FWS) says that wolves in the lower 48 states no longer face extinction (see *Nature* 501, 143–144; 2013). But the scientists, appointed by the FWS to review its policy, found that the agency failed to use the best available science to reach its conclusions. The FWS is now reviewing its

proposal and will reopen it for public comment for 45 days. The government is expected to make a final decision this year.

Growing fish farms

Some 62% of all the seafood eaten globally will be farmed by 2030, the World Bank predicts in a report released on 5 February. Fish farming contributed 40% of total supply in 2010, and is projected to expand to meet growing demand from regions including Asia, the bank says. Depleted wild fish stocks will also contribute to the growth in farmed fish, the report notes.

Climate hubs

US agriculture secretary Tom Vilsack announced the creation of seven regional 'climate hubs' on 5 February, to help farmers to reduce

carbon emissions and cope with climate change. The hubs will provide climate data and assessments, and will support agricultural research. They are part of a move by President Barack Obama's administration to advance climate science at regional and local levels.

UK funding cuts

English universities will lose an extra £125 million (US\$205 million) from their funding in 2014–15. The government's Department for Business, Innovation and Skills announced the cut from the budget of the Higher Education Funding Council for England (which will now stand at £4.1 billion) on 10 February. The annual grant letter setting out the allocations says that the council's core research funding

(£1.57 billion annually) will not fall. Cuts will hit teaching grants and funds targeted at helping the poorest students. See go.nature.com/lyb21y for more.

BUSINESS

Stem-cell ruling

Certain types of stem-cell treatment should be regulated as drugs, a US appeals court decided on 4 February. The ruling rejects claims by Regenerative Sciences of Broomfield, Colorado, that its stem-cell therapy, used in orthopaedic applications, is a medical procedure that does not require regulation by the Food and Drug Administration. The court decided that processing the cells for use in the therapy made the end product a drug. The company has not completed clinical trials on the treatment so cannot market the product in the United States. It does sell it in the Cayman Islands (see *Nature* **488**,14; 2012).

PEOPLE

Italian space head

The president of the Italian Space Agency resigned on 7 February following allegations of bribery and corruption. Enrico Saggese (**pictured**) is accused of accepting bribes in exchange



for contracts, inappropriately promoting friends and using public money to finance luxurious business travel, including a 3-week trip for 33 people to a rocket launch in Vandenberg, California. Saggese denies the charges. The agency has an annual budget of more than €550 million (US\$750 million) and is the third-largest contributor to the European Space Agency.

Insider trading

On 6 February, a US court found former hedge-fund manager Mathew Martoma guilty of leading a massive insider-trading scheme using confidential information about an Alzheimer's disease clinical trial. Sidney Gilman, a neurologist formerly at the University of Michigan in Ann Arbor, admitted to leaking data from the trial to Martoma before the results were made

public. Martoma's company netted US\$276 million in illicit gains, making it the largest insider-trading case ever charged by the US Securities and Exchange Commission. See go.nature.com/b8t9qx for more.

FACILITIES

Array revival

The US National Oceanic and Atmospheric Administration (NOAA) has decided to restore an ailing Pacific Ocean monitoring network used to observe and predict the periodic and disruptive ocean-warming event known as El Niño. The 2014 federal-government budget, approved last month, would provide US\$4 million — an increase of about one-third — to expand charter operations and revive the array, NOAA told *Nature* on 4 February. See go.nature.com/nrhvbo for more.

RESEARCH

Cancer genomics

Tumour genome sequencing may identify targeted treatments for only a fraction of patients with advanced breast cancer, according to research published on 7 February (F. André *et al.* *Lancet Oncol.* <http://doi.org/rdh>; 2014). Of 423 people studied, the authors identified only 13% with mutations

COMING UP

11 FEBRUARY

European member states will vote on whether to approve a genetically modified maize (corn) crop called Pioneer 1507, which is engineered to be resistant to insect pests.

12–13 FEBRUARY

London hosts a global conference on the illegal trade in wildlife. In May 2013, the United Nations officially characterized international wildlife and timber trafficking as a “serious crime”, which carries a minimum prison sentence of four years. See go.nature.com/htpd2l

that matched an available experimental treatment. The rate of successful therapy-matching falls short of claims from some tumour-sequencing companies, but researchers expect odds to improve as more drugs move into clinical trials.

Sharing drug data

Ten major pharmaceutical companies have agreed to share data from early-stage trials with each other and with academic researcher, as part of a US\$230-million venture with the US National Institutes of Health. The Accelerating Medicines Partnership, announced on 4 February, is designed to speed up identification of biomarkers and promising drug targets for four diseases: Alzheimer's, rheumatoid arthritis, lupus and diabetes. Once the data are shared, the companies will be free to pursue proprietary research towards their own therapies. See go.nature.com/rn12cl for more.

➔ **NATURE.COM**

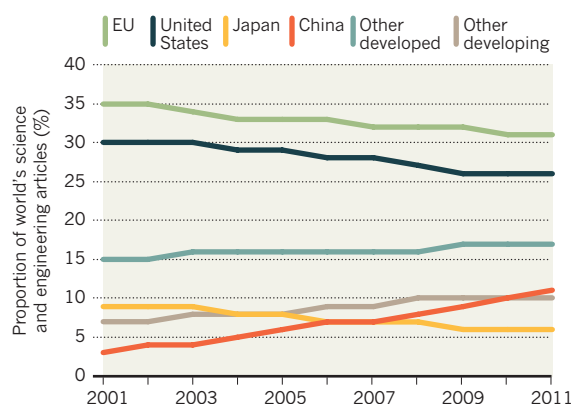
For daily news updates see: www.nature.com/news

TREND WATCH

In *Science and Engineering Indicators 2014*, a biennial round-up of global trends in research, the US National Science Foundation notes that the number of papers authored or co-authored by Chinese scientists grew by an average of more than 15% per year between 2001 and 2011 — leading the country to an 11% share of world output as the United States' share declined to 26%. The total world output of research papers grew by 2.8% per year. See go.nature.com/wfptv8 for more.

CHINA'S RESEARCH RISE

The US National Science Foundation charts the growing proportion of scientific articles with Chinese authors.



NEWS IN FOCUS

ECOLOGY Iconic wolf and moose study could be on its last legs **p.140**

ANTHROPOLOGY Delicate negotiations over remains of Clovis boy **p.142**

HEALTH CARE Medicine goes hyper-personal and hands-on in Seattle **p.144**



STATISTICS Is it time to take the P out of science? **p.150**

PUBLIC HEALTH ENGLAND



Research Agency in Stockholm, who took part in the research.

The IOR, a Home Office-led initiative, marks a shift in UK policy on dealing with CBR events, and the research on which it is based has caught the attention of international agencies. Until now, UK plans for large-scale incidents involved waiting for specialized teams to arrive wearing hazmat suits, and bringing portable showering tents and other equipment. But with fewer specialized teams than emergency services, delays would be inevitable — an important factor given that many chemical agents, such as sarin nerve gas, can be fast-acting. The new plan eases this problem by placing the initial response emphasis on those first at the scene.

The changes stem from the findings of a European programme called ORCHIDS (Optimisation through Research of Chemical Incident Decontamination Systems). Led by the UK Health Protection Agency (HPA), it involved Czech, French and Swedish security-research centres, as well as collaborators in Australia, Canada, Japan and the United States. A companion UK project, ORCHIDS-2, was carried out by the HPA and by researchers at the University of Hertfordshire in Hatfield and the University of Surrey in Guildford.

The ORCHIDS programmes ran in parallel from 2008 to 2011. Some of their findings have already been published, but much is only now being declassified, explains Robert Chilcott, a toxicologist at the University of Hertfordshire and a lead scientist on both programmes.

The aim was to improve the science base for emergency-response policies to civilian CBR attacks. Much existing research on decontamination following such attacks has been conducted by the military and is geared towards combat situations; it is often not adapted to mass civilian incidents, Chilcott says.

The ORCHIDS research quantified the effects of disrobing and decontamination protocols on the speed and effectiveness of the removal of agents. Various scenarios were tested using *in vitro* models and animal experiments before being replicated in humans.

The scientists found that clothes could provide significant initial protection, but that it was important to remove them quickly to prevent agents from penetrating through to the skin. They also found that dry decontamination was effective — this is crucial, because sufficient warm soapy water is unlikely to be immediately available. Combined disrobing ►

Specialist decontamination teams have until now been the first line of defence against bioterrorist attacks.

EMERGENCY RESPONSE

UK rolls out terror-attack plan

Front-line services being trained in new approach to dealing with decontamination of victims in direct aftermath of an event.

BY DECLAN BUTLER

A quick wipe-down with a paper towel can help to stem contamination of casualties in the event of a terrorist attack. That research finding is just one of several simple recommendations in the UK government's plan, now being rolled out, to treat mass civilian casualties in the immediate aftermath of a large-scale chemical, biological or radiological (CBR) attack, or an incident involving hazardous materials.

The Initial Operational Response (IOR)

programme, which is based on a raft of European research, largely classified until now, aims to help ambulance staff and other front-line emergency services to quickly deal with victims before specialists arrive. It introduces disrobing and improvised dry decontamination — patting down with paper towels or other readily available absorbent materials, for example — as the default initial response.

“Any decontamination has to be ‘quick and dirty’, to try to increase the chance of survival and to minimize spread of contamination,” says Ola Claesson of the Swedish Defence

► and improvised decontamination bought precious time, the research found.

The ORCHIDS project also investigated how showers could best be used once they became available. It tested various regimens, and found, for example, that simple measures such as flannel use and vigorous towel-drying greatly improved the removal of

“It gives us all credibility to go back to first responders and communities and say that this is scientifically based.”

agents. Showering for longer than 90 seconds provided little added benefit, a finding that could maximize the rate at which casualties are processed.

“Overall, the new IOR establishes the primary need to both disrobe and decontaminate individuals as soon as practically possible using any available method, and is based on experimental evidence,” says Chilcott. Training of emergency services in IOR techniques has already begun.

ORCHIDS data have been shared confidentially with security agencies in the United States and other countries. In 2012, the US Department of Health and Human Services, citing a need for similar research on US responses, awarded Chilcott a US\$6-million, two-year grant to apply ORCHIDS methodologies to optimize, for example, the mainstay of the US rapid on-scene response — the ‘ladder pipe’ system. In this, two fire trucks spray low-pressure water to form a decontamination corridor. The research is studying the effectiveness of the ladder-pipe system on clothed individuals.

US guidelines currently recommend that casualties disrobe to at least their underwear, but note that forcing people to undress could make them uncooperative and cause delays. The dilemma highlights the importance of such psychosocial aspects in the design of mass civilian decontamination plans (see go.nature.com/3jxvle).

Robert Gougelet, a researcher in emergency medicine at Dartmouth College in Hanover, New Hampshire, who is familiar with the ORCHIDS work, describes it as “novel and very credible”. He says that he has long advocated the need for a similar initial response phase in the United States. “One of the issues we have is that it’s easier for people just to sit back and say, ‘We are going to wait for the regional teams to arrive three hours later’. We are trying to change that way of thinking,” he adds.

Gougelet feels that having an improved solid science base will help. “It gives us all credibility to go back to the first responders and communities and say that this is scientifically based, this is the credible thing to do, the state of the art, and this is what you should be planning for.” ■



Wolves on Isle Royale in Lake Superior are suffering from spine problems as a result of inbreeding.

ECOLOGY

Iconic island study on its last legs

Ecologists call for genetic rescue of Isle Royale’s inbred wolves.

BY EMMA MARRIS

Since 1958, ecologists have watched wolf and moose populations on Isle Royale in Lake Superior wax and wane in response to each other, disease and the weather. But for the longest predator–prey study in the world, the wolf is now at the door. Devastated by inbreeding, the wolf population has dropped from 30 individuals a decade ago to just 10 spotted in field counts so far this year, leading the US National Park Service to consider importing new animals for a ‘genetic rescue’.

Now, nature is intervening — and could either save the landmark project without the need for tranquillizer darts and wolf crates, or sound its death knell. As temperatures plummeted last month, Lake Superior froze for the first time in six years. The 24-kilometre ice bridge could let wolves from the Canadian mainland cross to the US island, bringing an influx of genes (see ‘Wolf island’). But project scientists say that the opposite is more likely: free to roam, the last wolves could leave the island in search of mates.

That would put an end to a study that has

provided textbook ecology lessons for generations. It has shown how predation can structure populations of prey: when wolf numbers plummet, moose populations tend to soar (see ‘Ecosystem in flux’). And it has offered insights into wolf behaviour, moose physiology, the life cycle of moose ticks and how wolves might be driven to form packs to ward off scavengers such as ravens, rather than for any hunting advantage.

ROLF PETERSON/MINNEAPOLIS STAR TRIBUNE/ICIT/NEWS.COM



SOURCE: JOHN VUCETICH/ROLF PETERSON

Through the decades, the search for cause and effect in the ecosystem has been rendered much easier by isolation from the mainland's human and animal populations. Occasionally, however, Lake Superior freezes. The very first wolves came to Isle Royale over an ice bridge in the early 1940s, some 30 years after the first moose. The lake froze nearly every year at the beginning of the study, but that has changed. The most recent ice bridge was in 2008; before that, the last one was in 1997, when a wolf that biologists called 'the old grey guy' came to the island. He sired 34 pups and provided a rare boost of genes that doubled the population by the mid-2000s.

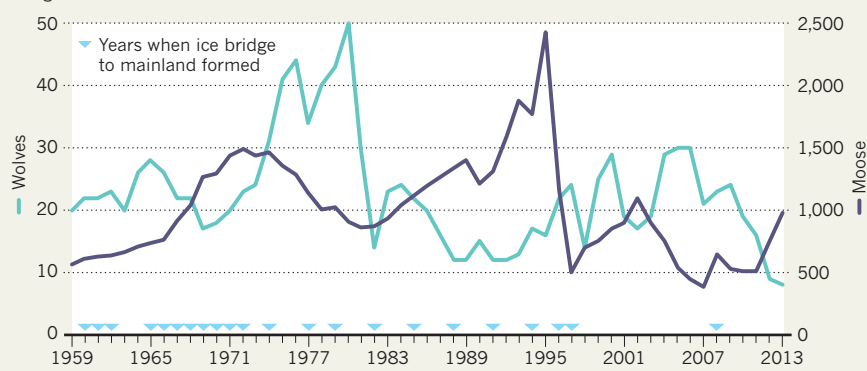
Whether any wolves have crossed this year's ice bridge will not be clear immediately. The scientists are conducting their annual population survey, and are flying along the island shore in their Super Cub plane two or three times a week, but snow fills wolf tracks very quickly. If new wolves do arrive, their presence will probably be confirmed in the coming months, when DNA is extracted from faeces samples.

John Vucetich, co-leader of the project and an ecologist at Michigan Technological University in Houghton, says that the need for an influx of genes is becoming urgent. In the past two decades, wolf skeletons have displayed spinal deformities that can painfully pinch nerves and affect gait and generally reduce fitness. According to work led by Vucetich and Rolf Peterson, also an ecologist at Michigan Technological University, this might explain why the number of moose needed to support a given number of wolves has increased: the predators' attacking efficiency may be compromised (J. Räikkönen *et al. Biol. Conserv.* **142**, 1025–1031; 2009).

For Vucetich, genetic rescue is required not so much to maintain the continuity of the study as to preserve the ecosystem. Moose eat balsam fir trees. When the moose population expands, unchecked by predation, fewer fir seedlings can grow large enough to 'escape' into the canopy above the reach of moose and reproduce. There is already a missing generation of trees from between about 1910, when the moose arrived on the island, and 1940, when the wolves came. Most of Isle Royale's balsam firs are thus either older than 100 years and near the end of their lives, or young and short enough to be browsed to death. If the

ECOSYSTEM IN FLUX

Moose and wolf populations on Isle Royale have oscillated in response to each other, the weather and disease, but inbreeding has now reduced the wolf population. An ice bridge in 2014 could allow new wolves to arrive.



trees do not achieve escape in the next decade or so, says Vucetich, "large portions of Isle Royale are not going to generate balsam fir, which is a really basic component of a boreal forest ecosystem".

Many scientists familiar with Isle Royale support genetic rescue, especially because human activity has contributed to the current population crash. Climate change has led to the decreasing frequency of ice bridges. Canine parvovirus, probably caught from a domestic dog, caused the wolf population to fall from around 50 to 14 in the early 1980s. And in 2012, three wolves were found dead in an abandoned mining pit. Given this history

A genetic rescue could set a precedent for intervention in other parks.

of human influence, the argument that leaving the wolves alone would be allowing nature to take its course does not sway most ecologists.

David Mech, a US Geological Survey wolf biologist based in St Paul, Minnesota, argues in favour of "watchful waiting". He says that much can be learned from studying how inbreeding affects population persistence, and that the knowledge would be useful for conservation biologists, who often need to nurture small, inbred populations of endangered species. He is not convinced that the wolves will die out; they have hit low numbers before and bounced back, he notes. And even if they do disappear, new wolves can be brought in quickly.

But Vucetich says that it could be five years before the last wolf dies and scientists confirm its demise, and another five before federal bureaucracies approve a genetic rescue and a pack develops into a predation force. He fears that a decade without significant moose predation would leave the fir trees devastated.

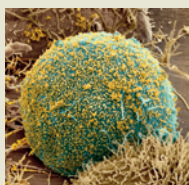
Phyllis Green, superintendent of the Isle Royale National Park, is considering three alternatives: doing nothing; watchful waiting followed by reintroduction if the population hits zero; or genetic rescue. She has not initiated a formal decision-making process, and will not commit to a timeline, but says that she wants to make a decision in consultation with her regional and national directors "before we run out of options".

She is proceeding cautiously, she says, in part because of the implications of her decision. The mandate of the National Parks Service, as enshrined in a 1916 Act of Congress, is to "conserve the scenery and the natural and historic objects and the wild life therein and to provide for the enjoyment of the same in such manner and by such means as will leave them unimpaired for the enjoyment of future generations". Generally, this has meant a hands-off approach, but a genetic rescue could set a precedent for interventions to counteract the effects of climate change in other parks.

Green knows that many scientists are in favour of genetic rescue, but she also hears from "wilderness-oriented" advocates who urge her not to intervene. "It is one of the wicked problems," she says. ■ [SEE EDITORIAL P. 132](#)

MORE ONLINE

TOP NEWS



Crowd-funded HIV trial accused of 'preying' on patients
go.nature.com/kujkd

MORE NEWS

- European Union tries to resolve mismatches with European Space Agency go.nature.com/pgm5ka
- Autism symptoms prevented in mice go.nature.com/m84wul
- Simulated blindness makes mice hear better go.nature.com/in212b

NATURE PODCAST



'Atomtronics' edge closer, an advance in nuclear fusion, and the evolution of colour nature.com/nature/podcast



Humans from the Clovis culture used characteristic stone points (brown) and rod-shaped bone tools.

ANTHROPOLOGY

Ancient genome stirs ethics debate

Sequencing of DNA from Native American ‘Clovis boy’ forces researchers to rethink handling of tribal remains.

BY EWEN CALLAWAY

The remains of a young boy, ceremonially buried some 12,600 years ago in Montana, have revealed the ancestry of one of the earliest populations in the Americas, known as the Clovis culture.

Published in this issue of *Nature*, the boy’s genome sequence shows that today’s indigenous groups spanning North and South America are all descended from a single population that trekked across the Bering land bridge from Asia (M. Rasmussen *et al. Nature* **506**, 225–229; 2014). The analysis also points to an early split between the ancestors of the Clovis people and a second group, whose DNA lives on in populations in Canada and Greenland (see page 162).

But the research underscores the ethical minefield of studying ancient Native American remains, and rekindles memories of a bruising legal fight over a different human skeleton in the 1990s.

To avoid such a controversy, Eske Willerslev, a palaeobiologist at the University of Copenhagen who led the latest study, attempted to involve Native American communities. And so he embarked on a tour of Montana’s Indian reservations last year, talking to community

members to explain his work and seek their support. “I didn’t want a situation where the first time they heard about this study was when it’s published,” he says.

Construction workers discovered the Clovis burial site on a private ranch near the small town of Wilsall in May 1968 (see ‘Ancient origins’). About 100 stone and bone artefacts, as well as bone fragments from a male child aged under two, were subsequently recovered.

The boy’s bones were found to date to the end of the Clovis culture, which flourished in the central and western United States between about 13,000 and 12,600 years ago. Carved elk bones

found with the boy’s remains were hundreds of years older, suggesting that they were heirlooms. The ranch, owned by Melvyn and Helen Anzick, is the only site yet discovered at which Clovis objects exist alongside human bones. Most of the artefacts now reside in a museum, but researchers returned the human remains to the Anzick family in the late 1990s.

At that time, the Anzicks’ daughter, Sarah, was conducting cancer and genome research at

the National Institutes of Health in Bethesda, Maryland, and thought about sequencing genetic material from the bones. But she was wary of stoking a similar debate to the one surrounding Kennewick Man, a human skeleton found on the banks of the Columbia River in Kennewick, Washington, in July 1996. Its discovery sparked an eight-year legal battle between Native American tribes, who claimed that they were culturally connected to the individual, and researchers, who said that the roughly 9,000-year-old remains pre-dated the tribes.

The US government sided with the tribes, citing the federal Native American Graves Protection and Repatriation Act (NAGPRA). The act requires that human remains discovered on federal lands — as Kennewick Man was — are returned to affiliated tribes for reburial. But a court ruled that the law did not apply, largely because of the age of the remains, and ordered that Kennewick Man be stored away from public view in a museum.

Sarah Anzick sought the advice of local tribes over the Clovis boy, but she could not reach a consensus with the tribes on what to do. She gave up on the idea, stored the bones in a safe location and got on with her other research.

In 2009, archaeologist Michael Waters, of Texas A&M University in College Station, contacted Anzick with the idea of sending the remains to Willerslev’s lab. (In early 2010, the lab published one of the first genome sequences of an ancient human, a 4,000-year-old resident of Greenland; see M. Rasmussen *et al. Nature* **463**, 757–762; 2010.) “I said, ‘I will allow you guys to do this, but I want to be involved,’” recalls Anzick, who has published more than a dozen papers in leading journals.

In Copenhagen, she extracted DNA from fragments of the boy’s skull ready for mitochondrial genome sequencing, which offers a snapshot of a person’s maternal ancestry. Back in Montana months later, she received the sequencing data and discovered that the genome’s closest match was to present-day Native Americans. “My heart just stopped,” she says.

RIGHT TO REMAINS

After Willerslev’s team confirmed the link by sequencing the boy’s nuclear genome (a more detailed indicator of ancestry), Willerslev sought advice from an agency that handles reburial issues. He was told that, because the remains were found on private land, NAGPRA did not apply and no consultation was needed. Despite this, Willerslev made his own attempt to consult local tribes. This led to a meeting in September at the burial site, with Anzick, Willerslev and their co-author Shane Doyle, who works in Native American studies at Montana State University in Bozeman, and is a member of the Crow tribe.

“That place is very special to me, that’s my ancestral homeland,” says Doyle. He

ROBERT L. WALKER

told Willerslev and Anzick that they should rebury the child where he was found. “I think you need to put the little boy back where his parents left him,” Doyle recalls telling them.

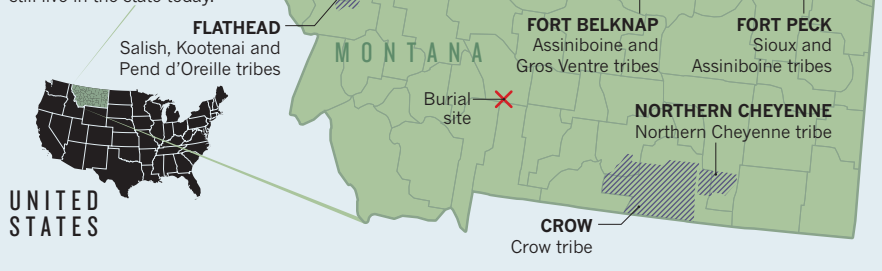
Doyle and Willerslev then set off on a 1,500-kilometre road trip to meet representatives of four Montana tribes; Doyle later consulted another five. Many of the people they talked to had few problems with the research, Doyle says, but some would have preferred to have been consulted before the study started, and not years after.

Willerslev says that researchers studying early American remains should assume that they are related to contemporary groups, and involve them as early as possible. But it is not always clear whom to contact, he adds, particularly when remains are related to groups spread across the Americas. “We have to engage with Native Americans, but how you deal with that question in practice is not an easy thing,” he says.

Hank Greeley, a legal scholar at Stanford University in California who is interested in the legal and ethical issues of human genetics, commends the approach of Willerslev’s team. But he says that there is no single solution to involving

ANCIENT ORIGINS

Montana yielded the remains of a child from the ancient Clovis culture, a population that is closely related to many Native Americans. Several tribes still live in the state today.



Native American communities in such research. “You’re looking to try to talk to the people who might be most invested in, or connected with, particular sets of remains,” he advises.

Dennis O’Rourke, a geneticist at the University of Utah in Salt Lake City, who studies ancient DNA from populations native to the islands around Alaska, notes that indigenous

groups have varying concerns: some want remains reburied, others do not, for instance.

The Montana tribes overwhelmingly wanted the Clovis boy’s bones interred. Plans for a reburial ceremony, possibly at an undisclosed site, are now being hashed out, with the Crow Nation playing a lead role. It is expected to take place in the spring, after the ground thaws. ■

PHARMACEUTICALS

NIH makes wary return to India

Some clinical trials funded by US agency resume, but strict regulations have put off others.

BY SARA REARDON

Allayed by pledges that India’s strict new rules for clinical trials will be eased, a few principal investigators funded by the US National Institutes of Health (NIH) are cautiously restarting studies there. Clarifications on the rules by the Drug Controller General of India (DCGI), and a promise to soften others, have allowed a small number of researchers to return to their work.

But for most others, the damage has already been done. Trial operators — both academic and industrial — have left India for other countries. Some researchers say that India’s clinical trials industry, which boomed over the past decade, may now be grinding to a halt.

“I don’t think we’re going to walk away from research in India, but it will certainly slow us down,” says Daniel Kuritzkes, a virologist at Harvard University in Cambridge, Massachusetts, and head of the international AIDS Clinical Trials Group (ACTG). Last March, after the new rules were announced, his group suspended two NIH-funded antiretroviral drug trials and a cervical cancer screening trial. The latter has resumed patient enrolment now that the DCGI has clarified that the rules apply only

to drug trials. But the antiretroviral trials were completed elsewhere, and the group has been unable to enrol Indian patients in new trials for tuberculosis and HIV drugs. “Obviously we have to continue research with or without our Indian colleagues,” Kuritzkes says. “We would much rather do it with them.”

Researchers had flocked to India because trials are cheap to conduct and there are many

people with disease who can be signed up. But a crackdown began in January last year, when India’s Supreme Court, concerned about allegations of unethical practices and deaths linked to trials (see go.nature.com/aa5xl2), imposed a ban on new ones. It told the DCGI to tighten its regulations, and the agency responded with a set of tough new rules to beef up patient protection.

In response, the NIH placed at least 35 ongoing clinical trials on hold. The law was vague and open to interpretation, researchers say: it seemed that trial sponsors would be required to provide medical care for trial participants for the rest of their lives, regardless of whether the trial itself had caused a medical problem. It also seemed that patients who received placebos, or for whom the drug did not work, would be entitled to compensation. “They went from one extreme to the other extreme,” says Kiran Mazumdar-Shaw, chief executive of Biocon, which is based in Bangalore and is one of India’s largest biotechnology companies.

Mazumdar-Shaw says that Biocon has already moved some of its trials to other countries, including ones that are far more expensive to operate in than India. She is particularly concerned about a requirement imposed by the Supreme Court in October to make a ▶



A woman is treated in a Mumbai cancer trial.

▶ videotape of each trial participant giving informed consent — for a vaccine trial, that could mean videotaping thousands of patients. Meanwhile, several other companies have closed their Indian branches. In October the clinical trials contractor Quintiles, based in Durham, North Carolina, closed its phase I unit in Hyderabad, and on 31 January AstraZeneca, headquartered in London, announced plans to shut its Bangalore centre, although

“The whole thing has had a very chilling effect for clinical trials and foreign sponsors.”

both companies maintain other facilities in India and say that the closures are due to restructuring. “The whole thing has had a very chilling effect for clinical trials and foreign sponsors,” says Amita Gupta, a researcher on ACTG trials at Johns Hopkins University in Baltimore, Maryland. “The motivation is coming from the right place, but the government overstepped.”

The health ministry that oversees the DCGI recognized this overstep and asked for recommended changes from a six-member panel led by Ranjit Roy Chaudhury, a pharmacologist and a former board member of the Medical Council of India. The panel’s July report gave various fixes, among which was a recommendation to scrap compensation for participants who received a placebo.

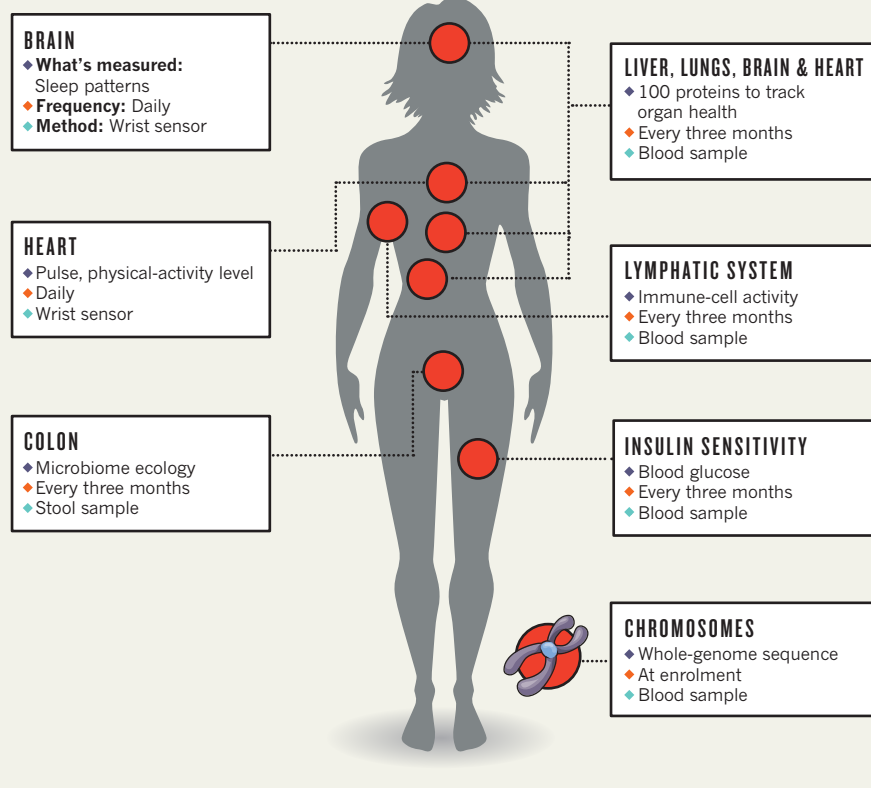
They also specified a new formula to calculate how much a participant who was harmed by a trial should be paid and suggested the formation of a national board to approve institutions’ ethics committees, which are responsible for determining whether an injury resulted from a trial and making sure informed consent was properly given. “We don’t want a single case of unethical practice”, says Y. K. Gupta, a pharmacologist at the All India Institute of Medical Sciences in New Delhi, who is helping to put together the board.

The DCGI has promised to implement most of the panel’s suggestions and has presented them to Parliament, which would need to codify the reforms. Already, some changes seem to be occurring; on 24 January, the agency announced that it would approve new clinical trials within six months of their submission. In addition, it clarified that an indemnity insurance requirement — insurance that would have been illegal for the NIH to buy under US law — can instead be purchased by partner institutions within India. Now that the rules are being clarified, Chaudhury hopes that researchers will reconsider their decision to suspend trials.

Yet, until researchers see laws in place, many will remain wary, says Amita Gupta. “I think a lot of groups don’t want to deal with the headache.” ■

AN EXAMINED LIFE

A nine-month study will collect data at daily and three-month intervals, and allow personalized interventions — such as changes in diet — as the study proceeds.



SOURCE: INSTITUTE FOR SYSTEMS BIOLOGY

HEALTH CARE

Medicine gets up close and personal

Long-term study will monitor healthy people in detail — and encourage them to respond to the results.

BY W. WAYT GIBBS

Leroy Hood, president of the Institute for Systems Biology (ISB) in Seattle, Washington, likes to talk about what he calls P4 medicine: health care that is predictive, preventive, personalized and participatory. Medicine today is a string of infrequent interventions prompted mainly by symptoms of illness. Hood argues instead for continuous management of health, making full use of whole-genome sequencing and biomarkers to correct disease before it gains a foothold.

In March, Hood will embark on the first big test of his ideas: a nine-month pilot study, dubbed the Hundred Person Wellness Project, in which 100 healthy individuals will be

intensively monitored (see ‘An examined life’), offered regular feedback and counselled on lifestyle changes such as shifts in their dietary or sleep habits. The effects of these behavioural changes on their health will, in turn, be tracked using a battery of diagnostic tests.

The study violates many rules of trial design: it dispenses with blinding and randomization, and will not even have a control group. But Hood is confident in its power to disrupt the conventional practice of medicine. “We hope to develop a whole series of stories about how actionable opportunities have changed the wellness of individuals, or have made them aware of how they can avoid disease,” he says.

If the pilot study works as hoped, it will

▶ videotape of each trial participant giving informed consent — for a vaccine trial, that could mean videotaping thousands of patients. Meanwhile, several other companies have closed their Indian branches. In October the clinical trials contractor Quintiles, based in Durham, North Carolina, closed its phase I unit in Hyderabad, and on 31 January AstraZeneca, headquartered in London, announced plans to shut its Banga-

“The whole thing has had a very chilling effect for clinical trials and foreign sponsors.”

lore centre, although both companies maintain other facilities in India and say that the closures are due to restructuring. “The whole thing has had a very chilling effect for clinical trials and foreign sponsors,” says Amita Gupta, a researcher on ACTG trials at Johns Hopkins University in Baltimore, Maryland. “The motivation is coming from the right place, but the government overstepped.”

The health ministry that oversees the DCGI recognized this overstep and asked for recommended changes from a six-member panel led by Ranjit Roy Chaudhury, a pharmacologist and a former board member of the Medical Council of India. The panel’s July report gave various fixes, among which was a recommendation to scrap compensation for participants who received a placebo.

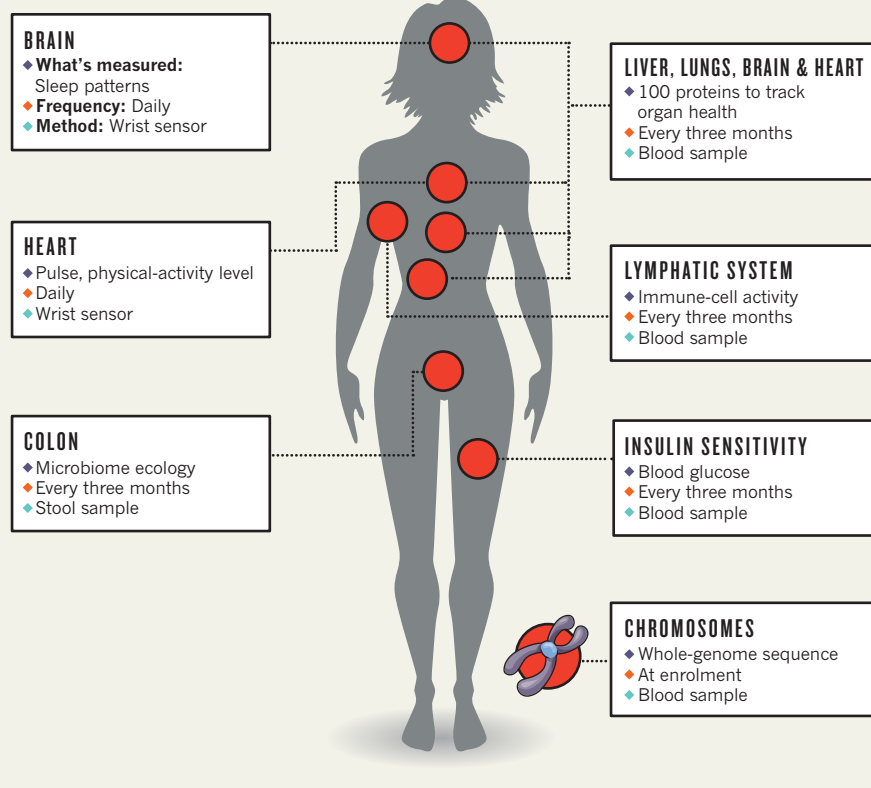
They also specified a new formula to calculate how much a participant who was harmed by a trial should be paid and suggested the formation of a national board to approve institutions’ ethics committees, which are responsible for determining whether an injury resulted from a trial and making sure informed consent was properly given. “We don’t want a single case of unethical practice”, says Y. K. Gupta, a pharmacologist at the All India Institute of Medical Sciences in New Delhi, who is helping to put together the board.

The DCGI has promised to implement most of the panel’s suggestions and has presented them to Parliament, which would need to codify the reforms. Already, some changes seem to be occurring; on 24 January, the agency announced that it would approve new clinical trials within six months of their submission. In addition, it clarified that an indemnity insurance requirement — insurance that would have been illegal for the NIH to buy under US law — can instead be purchased by partner institutions within India. Now that the rules are being clarified, Chaudhury hopes that researchers will reconsider their decision to suspend trials.

Yet, until researchers see laws in place, many will remain wary, says Amita Gupta. “I think a lot of groups don’t want to deal with the headache.” ■

AN EXAMINED LIFE

A nine-month study will collect data at daily and three-month intervals, and allow personalized interventions — such as changes in diet — as the study proceeds.



SOURCE: INSTITUTE FOR SYSTEMS BIOLOGY

HEALTH CARE

Medicine gets up close and personal

Long-term study will monitor healthy people in detail — and encourage them to respond to the results.

BY W. WAYT GIBBS

Leroy Hood, president of the Institute for Systems Biology (ISB) in Seattle, Washington, likes to talk about what he calls P4 medicine: health care that is predictive, preventive, personalized and participatory. Medicine today is a string of infrequent interventions prompted mainly by symptoms of illness. Hood argues instead for continuous management of health, making full use of whole-genome sequencing and biomarkers to correct disease before it gains a foothold.

In March, Hood will embark on the first big test of his ideas: a nine-month pilot study, dubbed the Hundred Person Wellness Project, in which 100 healthy individuals will be

intensively monitored (see ‘An examined life’), offered regular feedback and counselled on lifestyle changes such as shifts in their dietary or sleep habits. The effects of these behavioural changes on their health will, in turn, be tracked using a battery of diagnostic tests.

The study violates many rules of trial design: it dispenses with blinding and randomization, and will not even have a control group. But Hood is confident in its power to disrupt the conventional practice of medicine. “We hope to develop a whole series of stories about how actionable opportunities have changed the wellness of individuals, or have made them aware of how they can avoid disease,” he says.

If the pilot study works as hoped, it will

expand in several phases until it encompasses 100,000 subjects monitored over 25 years. The ISB is paying for the first hundred people through private donations and has budgeted around US\$10,000 per person. Hood expects those costs to drop drastically in a larger study, thanks to economies of scale and rapidly evolving diagnostic technologies. But he acknowledges the challenge of securing the hundreds of millions of dollars that a generation-long trial would require.

Even in its pilot phase, the project is unusually thorough. The ISB will sequence the whole genome of each participant at the outset. And in later phases, Hood says, the study team will also examine epigenetics: methylation and other modifications to DNA that can reflect environmental exposures. But that is just the tip of the data-collection iceberg.

Participants will be asked to wear digital devices that will continuously record their physical activity, heart rate and sleep patterns; subjects will periodically upload those data to the institute's systems. Every three months, researchers will gather samples of participants' blood, urine, saliva and stool. They will measure five biochemicals in saliva and urine, and sequence the stool samples to track the ecology of major microbial species in the gut. Blood-chemistry screens will extend well beyond the usual tests for cholesterol and glucose to include 20 less-commonly monitored variables, such as C-reactive protein — which signals inflammation at high levels. Hood's teams will also monitor about 100 organ-specific proteins that, he says, are sensitive markers for transitions from health to disease in mouse and cell models.

The point of the study — and of P4 medicine in general — is to detect those transitions and respond to them before symptoms appear. To that end, participants (mostly residents of the Seattle area, invited through social media) will have full access to their personal 'cloud' of data points. Some will have enough scientific training to dive into the literature and interpret their data themselves. But Hood expects most to rely on ISB-provided wellness coaches and their own physicians to interpret the results and recommend medical treatment or changes in diet or behaviour.

"The most amazing thing here is the repeated assessment every three months for some of these markers," says Graham Colditz, head of the public-health sciences division at Washington University School of Medicine in St Louis, Missouri. Colditz is a former principal investigator of the Nurses' Health Study, which has followed some 120,000 US nurses for 37 years through biennial questionnaires, plus a one-time collection of toenail clippings and two sets of blood samples from a subgroup.

But giving all of the ISB study participants feedback, and pushing them to make changes while the study is under way, risks thwarting the researchers' ability to draw conclusions about what worked and what did not,

some researchers say. And Colditz thinks that without control subjects who do not receive personalized coaching, and with so many complex data streams to analyse, Hood's team might have a hard time separating real signals from noise in the data.

Hood acknowledges the possible problems. But he counters that the existing clinical-trial system is "totally failing" because it cannot



"We hope to develop a whole series of stories about how actionable opportunities have changed the wellness of individuals."

Leroy Hood

control for peoples' different genetics and environments. The Hundred Person Wellness Project, he adds, recognizes that those differences are paramount, and treats each person as his or her own experiment. "We can follow things very carefully with each person and see how they respond," he says.

Atul Butte, head of systems medicine at Stanford University School of Medicine in California, is inclined to grant Hood the benefit of the doubt.

He says that an innovative study of this kind might indeed pick up the earliest hints of disease: "They may not end up proving that all these individuals benefit from the study in one particular way, but they may end up showing that all these individuals benefit from the study in their own individual ways."

Yet some study participants might have an experience similar to that of Larry Smarr, a computer scientist at the University of California, San Diego, and an early pioneer of the 'quantified self' movement. For 14 years, he has regularly monitored his biometrics — taking blood, DNA and, most recently, stool samples. He responded to the results by exercising more and radically overhauling his diet. His highly examined life delivered some impressive benefits — he lost 11 kilograms.

But Smarr says that it has also led him to discover alarming spikes in his immune activity, and a strange microbial ecology in his gut. "I had about 10% *E. coli*; the average healthy person has 0.06%," he says. Physicians told him he had late-onset Crohn's disease — something that he suspects might have been inflicted by his dietary changes.

Smarr says he would love to participate in the ISB study in the hope of getting his digestive system back on track; Hood says that there are a few spots left and that he is welcome. "I'm 65, so it's questionable whether I'll be around for another 25 years" to see the end of the long-term study, says Smarr. "On the other hand, watching me go bad could be a good thing to have in the database." ■



Evidence is mounting that medication for ADHD doesn't make a lasting difference to schoolwork or achievement.

Ben Harkless could not sit still. At home, the athletic ten-year-old preferred doing three activities at once: playing with his iPad, say, while watching television and rolling on an exercise ball. Sometimes he kicked the walls; other times, he literally bounced off them.

School was another story, however. Ben sat in class most days with his head down on his desk, "a defeated heap", remembers his mother, Suzanne Harkless, a social worker in Berkeley, California. His grades were poor, and his teacher was at a loss for what to do.

Harkless took Ben to a therapist who diagnosed him with attention deficit hyperactivity disorder (ADHD). He was prescribed methylphenidate, a stimulant used to improve focus in people with the condition.

Harkless was reluctant to medicate her child, so she gave him a dose

BY KATHERINE SHARPE

on a morning when she could visit the school to observe.

"He didn't whip through his work, but he finished his work," she says. "And then he went on and helped his classmate next to him. My jaw dropped."

ADHD diagnoses are rising rapidly around the world and especially in the United States, where 11% of children aged between 4 and 17 years old have been diagnosed with the disorder. Between half and two-thirds of those are put on medication, a decision often influenced by a child's difficulties at school. And there are numerous reports of adolescents and young adults without ADHD using the drugs as study aids.

As the drugs have become more widespread, so has their cultural cachet. Stimulant medications have gained a reputation for turbo-charging the intellect. Even news stories critical of their use refer to them as

“good-grade pills”, “cognitive enhancers” and “mental steroids”.

For most people with ADHD, these medications — typically formulations of methylphenidate or amphetamine — quickly calm them down and increase their ability to concentrate. Although these behavioural changes make the drugs useful, a growing body of evidence suggests that the benefits mainly stop there. Studies indicate that the improvements seen with medication do not translate into better academic achievement or even social adjustment in the long term: people who were medicated as children show no improvements in antisocial behaviour, substance abuse or arrest rates later in life, for example. And one recent study suggested that the medications could even harm some children¹.

After decades of study, it has become clear that the drugs are not as transformative as their marketers would have parents believe. “I don’t know of any evidence that’s consistent that shows that there’s any long-term benefit of taking the medication,” says James Swanson, a psychologist at the University of California, Irvine.

Now researchers are trying to understand why. The answer could lie in sub-optimal use of the drugs, or failure to address other factors that affect performance, such as learning disabilities. Or it could be that people place too much hope on a simple fix for a complex problem. “What we expect medication to do may be unrealistic,” says Lily Hechtman, a psychiatrist at McGill University in Montreal.

UNREALISTIC EXPECTATIONS?

In 1937, psychiatrist Charles Bradley noticed that problem children treated with a stimulant, benzedrine sulphate, became quieter, better behaved and more studious. Since then, studies have repeatedly demonstrated that stimulant medications reduce the core symptoms of ADHD, which include incessant, disruptive activity coupled with a lack of reflectiveness and inhibition. Stimulants work by increasing levels of the neurotransmitter dopamine in the brain, affecting regions involved in focus, self-control and the sense that an activity is rewarding. They take effect immediately, and they help as many as 80% of those with ADHD — one of the best response rates for a psychiatric drug.

Years of lab and classroom studies attest that the medications help affected children to perform in school. Treated children fidget less. They do better on laboratory tests requiring concentration and short-term memory. And they take better notes and hand in more homework, making fewer careless mistakes. Nora Volkow, director of the National Institute on Drug Abuse in Bethesda, Maryland, says that these benefits carry over into the real world, at least in the short term. “They help you pay attention,” she says. “The grades do improve.”

But the few studies that have examined the effects of ADHD medication much beyond a year have found that the benefits either vanish or shrink to clinically meaningless proportions.

In the early 1990s, as rates of stimulant prescriptions were beginning to climb, the National Institute of Mental Health in Bethesda, Maryland, funded a study to compare different treatments for the disorder. Known as the Multimodal Treatment Study of Children with ADHD, or MTA, the study randomized 579 children aged between seven and ten with ADHD to receive one of four treatments: stimulant medication, behaviour therapy, medication and behaviour therapy combined or whatever care they had already been receiving.

After 14 months, the groups treated with medication alone and medication plus behaviour therapy showed greater improvements in core ADHD symptoms than the other two groups. For academic achievement, only the group receiving medication and behaviour therapy combined outperformed the group receiving regular care². By three years in, the four groups had become indistinguishable on every measure³.

Treatment conferred no lasting benefit in terms of grades, test scores or social adjustment. Eight years later, it was the same story⁴. “Nothing we did could tease out and say there’s a long-term effect,” says Swanson, who was one of the lead investigators on the MTA.

The MTA’s findings are borne out in most of the studies that followed students for long periods of time. A literature review in 2012, which included studies that tracked children with ADHD for three years or more, found little evidence for a significant effect on standardised-test scores, grades or on the likelihood that a student would be held back a year⁵. A 2013 review of randomized controlled trials longer than 12 months similarly concluded that there is scant evidence for improvements in ADHD symptoms or academic performance lasting much beyond a year⁶.

There is even some evidence that ADHD medication could worsen outcomes. In 2013, a team of economists published a study¹ examining the effects of a policy change in Quebec that resulted in thousands of children being given prescriptions for

methylphenidate. The authors found that children who began taking it actually did worse at school and were more likely to drop out than those with similar levels of symptoms who did not receive drugs. Girls taking the drug had more emotional problems, and both sexes reported worse relationships with their parents.

There are a few studies that do show long-term gains in academic performance, but the boost is not large. A study that tracked 594 students aged 5–11 with ADHD found that those using medication for at least a year scored 3 points out of 100 higher on standardized maths tests and 5 points higher on reading tests than those not taking medication⁷. But this was not enough to close the test-score gap between those with ADHD and those without. And the gains faded over time even if the children stayed on the drugs, according to study co-author Stephen Hinshaw, a psychologist at the University of California, Berkeley.

In 2012, a study in Iceland — the only country where rates of stimulant medication use are comparable to those of the United States — found that although the scores of all children with ADHD declined, on average, on standardized maths tests between the ages of 9 and 12, those of students who started medication earlier during that period declined less than those who waited longer to start⁸.

It is possible that there are long-term benefits that studies so far have not captured. But given the abundance and consistency of the data, the drugs cannot be doing much for most of the millions of children who take them, says Alan Sroufe, a psychologist emeritus at the University of Minnesota in Minneapolis. “If they were, it wouldn’t be hard to detect.”

PUZZLING PARADOX

Researchers are beginning to address this paradox. How can medication that makes children sit still and pay attention not lead to better grades?

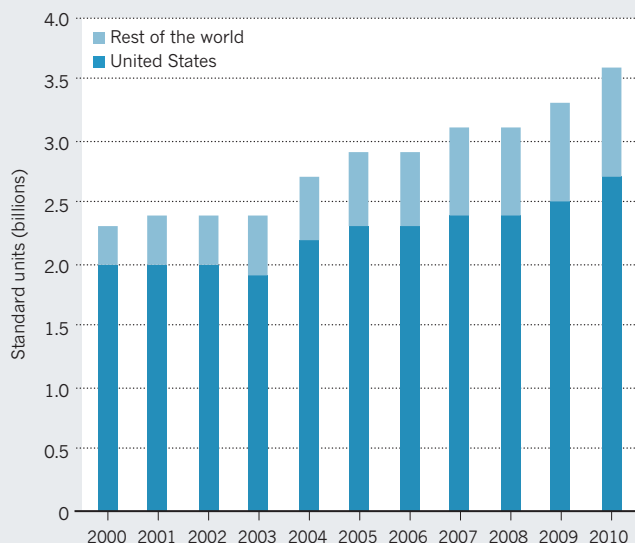
One possibility is that children develop tolerance to the drug. Dosage could also play a part: as children grow and put on weight, medication has to be adjusted to keep up, which does not always happen. And many children simply stop taking the drugs, especially in adolescence, when they may begin to feel that it affects their personalities. Children may also stop treatment because of side effects, which can include difficulty sleeping, loss of appetite and mood swings, as well as elevated heart rate.

Or it could be that stimulant medications mainly improve behaviour, not intellectual functioning. In the 1970s, two researchers, Russell Barkley and Charles Cunningham, noted that when children with ADHD took stimulants, parents and teachers rated their academic performance as vastly improved⁹. But objective measurements showed that the quality of their work hadn’t changed. What looked like achievement was actually manageability in the classroom. If medication made struggling

“ONLY ONE IN FOUR KIDS ARE GETTING ANYTHING CLOSE TO WHAT WE WOULD SAY IS GOOD TREATMENT.”

POPULAR PRESCRIPTIONS

The United States prescribes many more drugs for ADHD than the rest of the world put together, but other countries are beginning to narrow the gap.



children appear to be doing fine, they might be passed over for needed help, the authors suggested. Janet Currie, an economist at Princeton University in New Jersey, says that she might have been observing just such a phenomenon in the Quebec study that found lower achievement among medicated students¹.

And it may simply be that drugs are not enough. Stimulant medications have two core effects: they help people to sustain mental effort, and they make boring, repetitive tasks seem more interesting. Those properties help with many school assignments, but not all of them. Children treated with stimulants would be able to complete a worksheet of simple maths problems faster and more accurately than usual, explains Nora Volkow. But where flexibility of thought is required — for example, if each problem on a worksheet demands a different kind of solution — stimulants do not help.

BEYOND BELIEF

In people without ADHD, such as students who take the drugs without a prescription to help with school work, the intellectual impact of stimulants also remains unimpressive. In a 2012 study of the effects of the amphetamine Adderall on people without ADHD, psychologists at the University of Pennsylvania in Philadelphia found no consistent improvement on numerous measures of cognition, even though people taking the medication believed that their performance had been enhanced¹⁰.

Increased focus has benefits, say some experts, but many children with ADHD need help in more areas if they are to succeed at school. “Many things go into grades,” says Joshua Langberg, a psychologist at Virginia Commonwealth University in Richmond. “One of those is certainly a child’s behaviour and ability to focus, which medication does a nice job of improving. But they also include a child’s basic abilities in math and reading, their IQ and their ability to manage time and plan. It’s not clear why we would expect medication to impact those things.”

Some researchers think that the lack of evidence for long-term academic benefits is a result of flawed study design. Peter Jensen, a leader on the MTA study, says he believes that if the children had been maintained on the study’s protocol, the initial gains they made would have lasted. Longer randomly controlled trials would be challenging both from a technical and ethical standpoint, but the suggestion highlights another problem, namely the discrepancy between the optimal care given during a trial and that which most children receive.

After the 14-month, randomized trial period, participants in the MTA study began to receive what Jensen calls treatment ‘in the community’. He says it is typically of low quality. Few doctors monitor children closely enough to arrive at optimal dosage or identify and treat co-occurring conditions — such as depression and anxiety — that affect up to 70% of children with ADHD. “Only one in four kids are getting anything close to what we would say is good treatment,” Jensen says.

When the MTA team examined the follow-up data, it found that many non-medical factors play a big part in whether improvements last. The best predictor of a child’s response to treatment wasn’t which treatment they were assigned, but a cluster of factors that were present at the start. Children with more advantages — higher intelligence, better social skills, intact families, higher parental education, fewer conduct problems or higher socioeconomic status — were likely to make big strides and hold onto them no matter what the treatment was, whereas children without these advantages typically progressed more slowly and regressed after treatment stopped^{2–4}.

But disadvantaged children benefited when they received both medication and behaviour therapy. “The kids with the most problems needed the combination,” says Jensen, who adds that parents should have easier access to proven behaviour therapies. The effects of behavioural treatment don’t seem to be longer-lasting than those of medication, however: once active treatment stops, they dissipate.

Future studies might explore whether medication offers subtle benefits that are not reflected in test scores or grades. Many researchers think that a stint on medication, when it is needed, can create an upward spiral of self-esteem that may make a crucial difference to a child’s life — but there are no hard data to support this. “It may be that treatment doesn’t translate into better grades” in the long term, Volkow says. “But what I’d like to see is, are those kids overall better integrated?”

Some experts think that the focus on academic achievement is misguided — that the point of the drugs has never been to improve children’s grades, or increase their chances of admission to the best universities. “Medications are given for their short-term effects,” says Swanson. “Don’t expect medication to get rid of every problem a child has. But if the problem right now is not passing the second grade, or not having any friends in the third grade, we can do something about that now.”

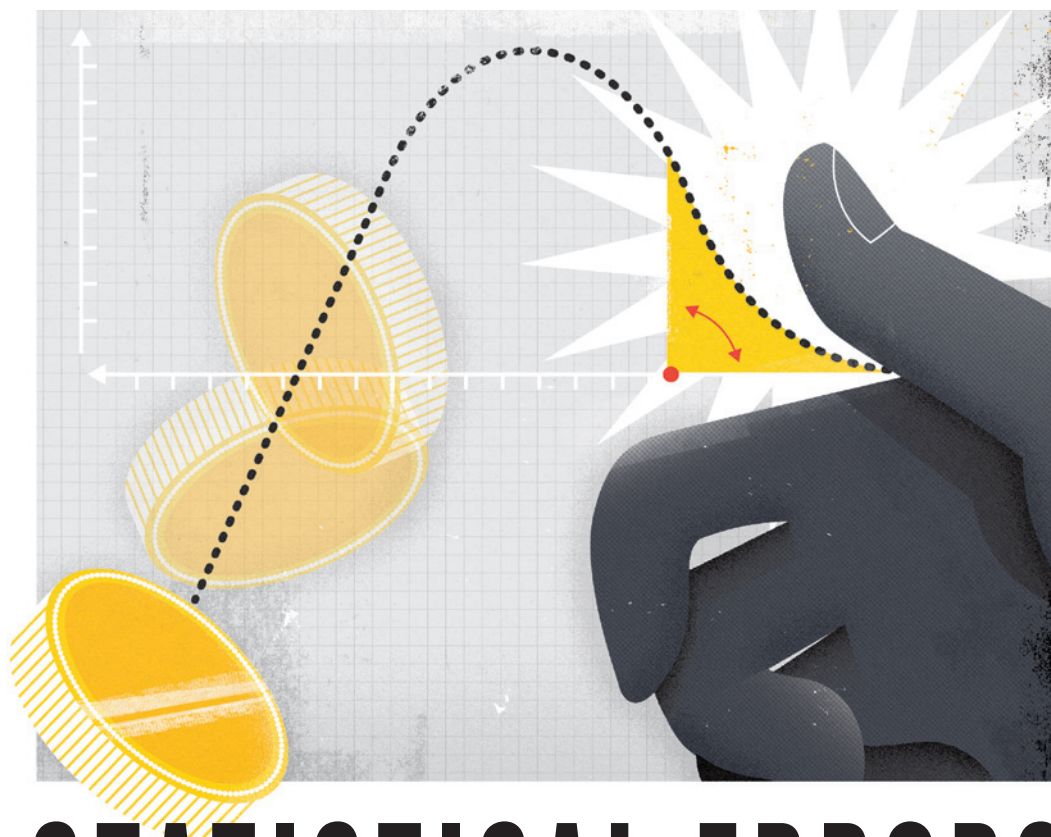
Some parents seem to understand that. Suzanne Harkless says that her hopes for medication are modest. She wants to keep Ben engaged in the fifth grade while she looks for a middle school that might provide him with the structure he needs. “My goal right now is not to get him into a good college,” she says. “My goal is to keep him in school.”

Other parents may pin unrealistic hopes on these drugs as their use goes up around the world (see ‘Popular prescriptions’). “Competition in today’s global economy is fuelling the dramatic increase in the use of ADHD medications, especially in the United States,” says Richard Scheffler, a health economist at the University of California, Berkeley, and co-author of a forthcoming book with Hinshaw on the growing popularity of ADHD drugs.

For Currie, the question comes down to transparency. “Parents do care about how their children are doing in school,” she says. “It’s misleading to tell parents that this will help their children succeed, when there’s no evidence that it’s the case.” ■

Katherine Sharpe is a science writer in Berkeley, California.

- Currie, J., Stabile, M. & Jones, L. E. *National Bureau of Economic Research Working Paper 19105* (NBER, 2013); available at <http://www.nber.org/papers/w19105>
- MTA Cooperative Group *Arch. Gen. Psychiatry* **56**, 1073–1086 (1999).
- Jensen, P. S. et al. *J. Am. Acad. Child Adolesc. Psychiatry* **46**, 989–1002 (2007).
- Molina, B. S. G. et al. *J. Am. Acad. Child Adolesc. Psychiatry* **48**, 484–500 (2009).
- Langberg, J. M. & Becker, S. P. *Clin. Child Fam. Psychol. Rev.* **15**, 215–233 (2012).
- Parker, J. et al. *Psychol. Res. Behav. Manag.* **6**, 87–99 (2013).
- Scheffler, R. M. et al. *Pediatrics* **123**, 1273–1279 (2009).
- Zoëga, H. et al. *Pediatrics* **130**, e53–e62 (2012).
- Barkley, R. A. & Cunningham, C. E. *Clin. Pediatr.* **17**, 85–92 (1978).
- Iliev, I., Boland, J. & Farah, M. J. *Neuropharmacology* **64**, 496–505 (2013).



STATISTICAL ERRORS

P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

For a brief moment in 2010, Matt Motyl was on the brink of scientific glory: he had discovered that extremists quite literally see the world in black and white.

The results were “plain as day”, recalls Motyl, a psychology PhD student at the University of Virginia in Charlottesville. Data from a study of nearly 2,000 people seemed to show that political moderates saw shades of grey more accurately than did either left-wing or right-wing extremists. “The hypothesis was sexy,” he says, “and the data provided clear support.” The *P* value, a common index for the strength of evidence, was 0.01 — usually interpreted as ‘very significant’. Publication in a high-impact journal seemed within Motyl’s grasp.

But then reality intervened. Sensitive to controversies over reproducibility, Motyl and his adviser, Brian Nosek, decided to replicate the study. With extra data, the *P* value came out as 0.59 — not even close to the conventional level of significance, 0.05. The effect had disappeared, and with it, Motyl’s dreams of youthful fame¹.

It turned out that the problem was not in the data or in Motyl’s analyses. It lay in the surprisingly slippery nature of the *P* value, which is neither as reliable nor as objective as most scientists assume. “*P* values are not doing their job, because they can’t,” says Stephen Ziliak, an economist at Roosevelt University in Chicago, Illinois, and a frequent critic of the way statistics are used.

For many scientists, this is especially worrying in light of the reproducibility concerns. In 2005, epidemiologist John Ioannidis of Stanford University in California suggested that most published findings are false²; since then, a string of high-profile replication problems has forced scientists to rethink how they evaluate results.

At the same time, statisticians are looking for better ways of thinking about data, to help scientists to avoid missing important information or acting on false alarms. “Change your statistical philosophy and all of a sudden different things become important,” says Steven

Goodman, a physician and statistician at Stanford. “Then ‘laws’ handed down from God are no longer handed down from God. They’re actually handed down to us by ourselves, through the methodology we adopt.”

OUT OF CONTEXT

P values have always had critics. In their almost nine decades of existence, they have been likened to mosquitoes (annoying and impossible to swat away), the emperor’s new clothes (fraught with obvious problems that everyone ignores) and the tool of a “sterile intellectual rake” who ravishes science but leaves it with no progeny³. One researcher suggested rechristening the methodology “statistical hypothesis inference testing”³, presumably for the acronym it would yield.

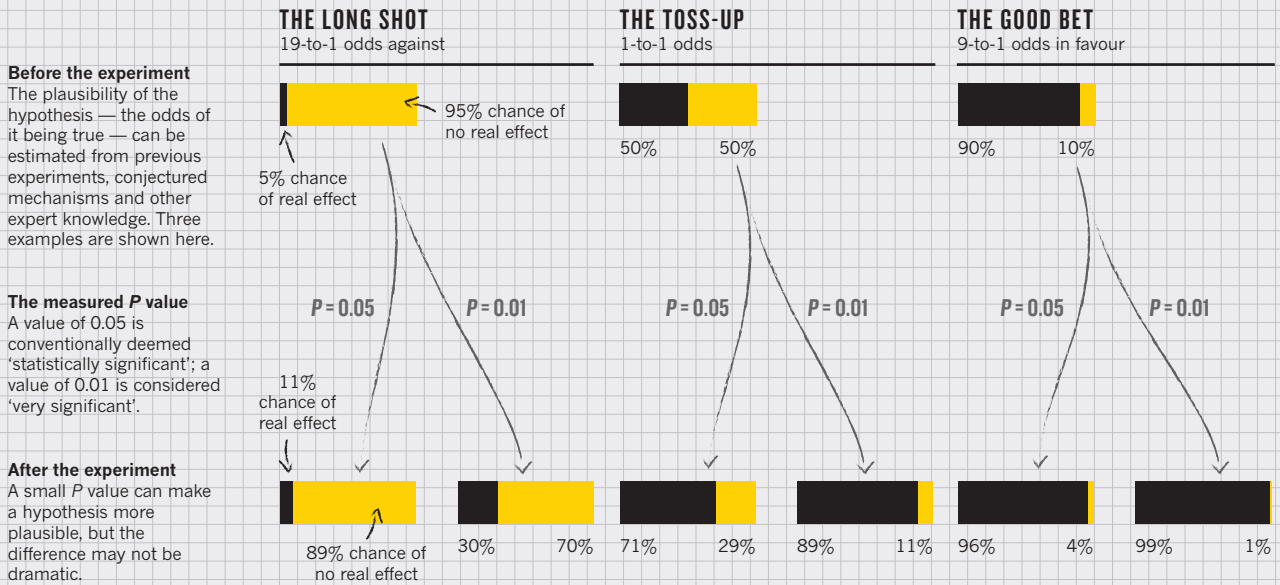
The irony is that when UK statistician Ronald Fisher introduced the *P* value in the 1920s, he did not mean it to be a definitive test. He intended it simply as an informal way to judge whether evidence was significant in the

DALE EDWIN MURRAY

PROBABLE CAUSE

A P value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

■ Chance of real effect
■ Chance of no real effect



old-fashioned sense: worthy of a second look. The idea was to run an experiment, then see if the results were consistent with what random chance might produce. Researchers would first set up a 'null hypothesis' that they wanted to disprove, such as there being no correlation or no difference between two groups. Next, they would play the devil's advocate and, assuming that this null hypothesis was in fact true, calculate the chances of getting results at least as extreme as what was actually observed. This probability was the P value. The smaller it was, suggested Fisher, the greater the likelihood that the straw-man null hypothesis was false.

For all the P value's apparent precision, Fisher intended it to be just one part of a fluid, non-numerical process that blended data and background knowledge to lead to scientific conclusions. But it soon got swept into a movement to make evidence-based decision-making as rigorous and objective as possible. This movement was spearheaded in the late 1920s by Fisher's bitter rivals, Polish mathematician Jerzy Neyman and UK statistician Egon Pearson, who introduced an alternative framework for data analysis that included statistical power, false positives, false negatives and many other concepts now familiar from introductory statistics classes. They pointedly left out the P value.

But while the rivals feuded — Neyman called some of Fisher's work mathematically "worse than useless"; Fisher called Neyman's approach "childish" and "horrible [for] intellectual freedom in the west" — other researchers lost patience and began to write statistics manuals for working scientists. And because

many of the authors were non-statisticians without a thorough understanding of either approach, they created a hybrid system that crammed Fisher's easy-to-calculate P value into Neyman and Pearson's reassuringly rigorous rule-based system. This is when a P value of 0.05 became enshrined as 'statistically significant', for example. "The P value was never meant to be used the way it's used today," says Goodman.

WHAT DOES IT ALL MEAN?

One result is an abundance of confusion about what the P value means⁴. Consider Motyl's study about political extremists. Most scientists would look at his original P value of 0.01 and say that there was just a 1% chance of his result being a false alarm. But they would be wrong. The P value cannot say this: all it can do is summarize the data assuming a specific null hypothesis. It cannot work backwards and make statements about the underlying reality. That requires another piece of information: the odds that a real effect was there in the first place. To ignore this would be like waking up with a headache and concluding that you have a rare brain tumour — possible, but so unlikely that it requires a lot more evidence to supersede an everyday explanation such as an allergic reaction. The more implausible the hypothesis — telepathy, aliens, homeopathy — the greater the chance that an exciting finding is a false alarm, no matter what the P value is.

➔ **NATURE.COM**
For more on statistics, see:
go.nature.com/xlj9lr

These are sticky concepts, but some statisticians have tried to

provide general rule-of-thumb conversions (see 'Probable cause'). According to one widely used calculation⁵, a P value of 0.01 corresponds to a false-alarm probability of at least 11%, depending on the underlying probability that there is a true effect; a P value of 0.05 raises that chance to at least 29%. So Motyl's finding had a greater than one in ten chance of being a false alarm. Likewise, the probability of replicating his original result was not 99%, as most would assume, but something closer to 73% — or only 50%, if he wanted another 'very significant' result^{6,7}. In other words, his inability to replicate the result was about as surprising as if he had called heads on a coin toss and it had come up tails.

Critics also bemoan the way that P values can encourage muddled thinking. A prime example is their tendency to deflect attention from the actual size of an effect. Last year, for example, a study of more than 19,000 people showed⁸ that those who meet their spouses online are less likely to divorce ($p < 0.002$) and more likely to have high marital satisfaction ($p < 0.001$) than those who meet offline (see *Nature* <http://doi.org/rcg>; 2013). That might have sounded impressive, but the effects were actually tiny: meeting online nudged the divorce rate from 7.67% down to 5.96%, and barely budged happiness from 5.48 to 5.64 on a 7-point scale. To pounce on tiny P values and ignore the larger question is to fall prey to the "seductive certainty of significance", says Geoff Cumming, an emeritus psychologist at La Trobe University in Melbourne, Australia. But significance is no indicator of practical relevance, he says: "We should be asking,

'How much of an effect is there?', not 'Is there an effect?'"

Perhaps the worst fallacy is the kind of self-deception for which psychologist Uri Simonsohn of the University of Pennsylvania and his colleagues have popularized the term *P*-hacking; it is also known as data-dredging, snooping, fishing, significance-chasing and double-dipping. "*P*-hacking," says Simonsohn, "is trying multiple things until you get the desired result" — even unconsciously. It may be the first statistical term to rate a definition in the online Urban Dictionary, where the usage examples are telling: "That finding seems to have been obtained through *p*-hacking, the authors dropped one of the conditions so that the overall *p*-value would be less than .05", and "She is a *p*-hacker, she always monitors data while it is being collected."

Such practices have the effect of turning discoveries from exploratory studies — which should be treated with scepticism — into what look like sound confirmations but vanish on replication. Simonsohn's simulations have shown⁹ that changes in a few data-analysis decisions can increase the false-positive rate in a single study to 60%. *P*-hacking is especially likely, he says, in today's environment of studies that chase small effects hidden in noisy data. It is tough to pin down how widespread the problem is, but Simonsohn has the sense that it is serious. In an analysis¹⁰, he found evidence that many published psychology papers report *P* values that cluster suspiciously around 0.05, just as would be expected if researchers fished for significant *P* values until they found one.

NUMBERS GAME

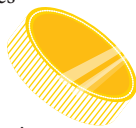
Despite the criticisms, reform has been slow. "The basic framework of statistics has been virtually unchanged since Fisher, Neyman and Pearson introduced it," says Goodman. John Campbell, a psychologist now at the University of Minnesota in Minneapolis, bemoaned the issue in 1982, when he was editor of the *Journal of Applied Psychology*: "It is almost impossible to drag authors away from their *p*-values, and the more zeroes after the decimal point, the harder people cling to them"¹¹. In 1989, when Kenneth Rothman of Boston University in Massachusetts started the journal *Epidemiology*, he did his best to discourage *P* values in its pages. But he left the journal in 2001, and *P* values have since made a resurgence.

Ioannidis is currently mining the PubMed database for insights into how authors across many fields are using *P* values and other statistical evidence. "A cursory look at a sample of recently published papers," he says, "is convincing that *P* values are still very, very popular."

Any reform would need to sweep through an entrenched culture. It would have to change

how statistics is taught, how data analysis is done and how results are reported and interpreted. But at least researchers are admitting that they have a problem, says Goodman. "The wake-up call is that so many of our published findings are not true." Work by researchers such as Ioannidis shows the link between theoretical statistical complaints and actual difficulties, says Goodman. "The problems that statisticians have predicted are exactly what we're now seeing. We just don't yet have all the fixes."

"THE *P* VALUE WAS NEVER MEANT TO BE USED THE WAY IT'S USED TODAY."



Statisticians have pointed to a number of measures that might help. To avoid the trap of thinking about results as significant or not significant, for example, Cumming thinks that researchers should always report effect sizes and confidence intervals. These convey what a *P* value does not: the magnitude and relative importance of an effect.

Many statisticians also advocate replacing the *P* value with methods that take advantage of Bayes' rule: an eighteenth-century theorem that describes how to think about probability as the plausibility of an outcome, rather than as the potential frequency of that outcome. This entails a certain subjectivity — something that the statistical pioneers were trying to avoid. But the Bayesian framework makes it comparatively easy for observers to incorporate what they know about the world into their conclusions, and to calculate how probabilities change as new evidence arises.

Others argue for a more ecumenical approach, encouraging researchers to try multiple methods on the same data set. Stephen Senn, a statistician at the Centre for Public Health Research in Luxembourg City, likens this to using a floor-cleaning robot that cannot find its own way out of a corner: any data-analysis method will eventually hit a wall, and some common sense will be needed to get the process moving again. If the various methods come up with different answers, he says, "that's a suggestion to be more creative and try to find out why", which should lead to a better understanding of the underlying reality.

Simonsohn argues that one of the strongest protections for scientists is to admit everything. He encourages authors to brand their papers '*P*-certified, not *P*-hacked' by including the words: "We report how we determined our sample size, all data exclusions (if any), all manipulations and all measures

in the study." This disclosure will, he hopes, discourage *P*-hacking, or at least alert readers to any shenanigans and allow them to judge accordingly.

A related idea that is garnering attention is two-stage analysis, or 'preregistered replication', says political scientist and statistician Andrew Gelman of Columbia University in New York City. In this approach, exploratory and confirmatory analyses are approached differently and clearly labelled. Instead of doing four separate small studies and reporting the results in one paper, for instance, researchers would first do two small exploratory studies and gather potentially interesting findings without worrying too much about false alarms. Then, on the basis of these results, the authors would decide exactly how they planned to confirm the findings, and would publicly preregister their intentions in a database such as the Open Science Framework (<https://osf.io>). They would then conduct the replication studies and publish the results alongside those of the exploratory studies. This approach allows for freedom and flexibility in analyses, says Gelman, while providing enough rigour to reduce the number of false alarms being published.

More broadly, researchers need to realize the limits of conventional statistics, Goodman says. They should instead bring into their analysis elements of scientific judgement about the plausibility of a hypothesis and study limitations that are normally banished to the discussion section: results of identical or similar experiments, proposed mechanisms, clinical knowledge and so on. Statistician Richard Royall of Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, said that there are three questions a scientist might want to ask after a study: 'What is the evidence?' 'What should I believe?' and 'What should I do?' One method cannot answer all these questions, Goodman says: "The numbers are where the scientific discussion should start, not end." ■ [SEE EDITORIAL P.131](#)

Regina Nuzzo is a freelance writer and an associate professor of statistics at Gallaudet University in Washington DC.

1. Nosek, B. A., Spies, J. R. & Motyl, M. *Perspect. Psychol. Sci.* **7**, 615–631 (2012).
2. Ioannidis, J. P. A. *PLoS Med.* **2**, e124 (2005).
3. Lambdin, C. *Theory Psychol.* **22**, 67–90 (2012).
4. Goodman, S. N. *Ann. Internal Med.* **130**, 995–1004 (1999).
5. Goodman, S. N. *Epidemiology* **12**, 295–297 (2001).
6. Goodman, S. N. *Stat. Med.* **11**, 875–879 (1992).
7. Gorroochurn, P., Hodge, S. E., Heiman, G. A., Durner, M. & Greenberg, D. A. *Genet. Med.* **9**, 325–321 (2007).
8. Cacioppo, J. T., Cacioppo, S., Gonzagab, G. C., Ogburn, E. L. & VanderWeele, T. J. *Proc. Natl Acad. Sci. USA* **110**, 10135–10140 (2013).
9. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psychol. Sci.* **22**, 1359–1366 (2011).
10. Simonsohn, U., Nelson, L. D. & Simmons, J. P. *J. Exp. Psychol.* <http://dx.doi.org/10.1037/a0033242> (2013).
11. Campbell, J. P. *J. Appl. Psych.* **67**, 691–700 (1982).

COMMENT

BIODIVERSITY A study of Earth's six bouts of extinctions **p.156**

SOUND Exploring the acoustics of prehistoric caves **p.158**

AUSTERITY Swingeing cuts threaten future of research in Portugal **p.159**

LAB LIFE More courses needed to teach scientists leadership skills **p.159**



FRANS LANTING/NATIONAL GEOGRAPHIC CREATIVE



Lowland rainforest in Manu National Park, Peru.

Managing forests in uncertain times

Increasing both forest stocks and timber harvest will buy time while we learn more about how trees absorb carbon, say **Valentin Bellassen** and **Sebastiaan Luyssaert**.

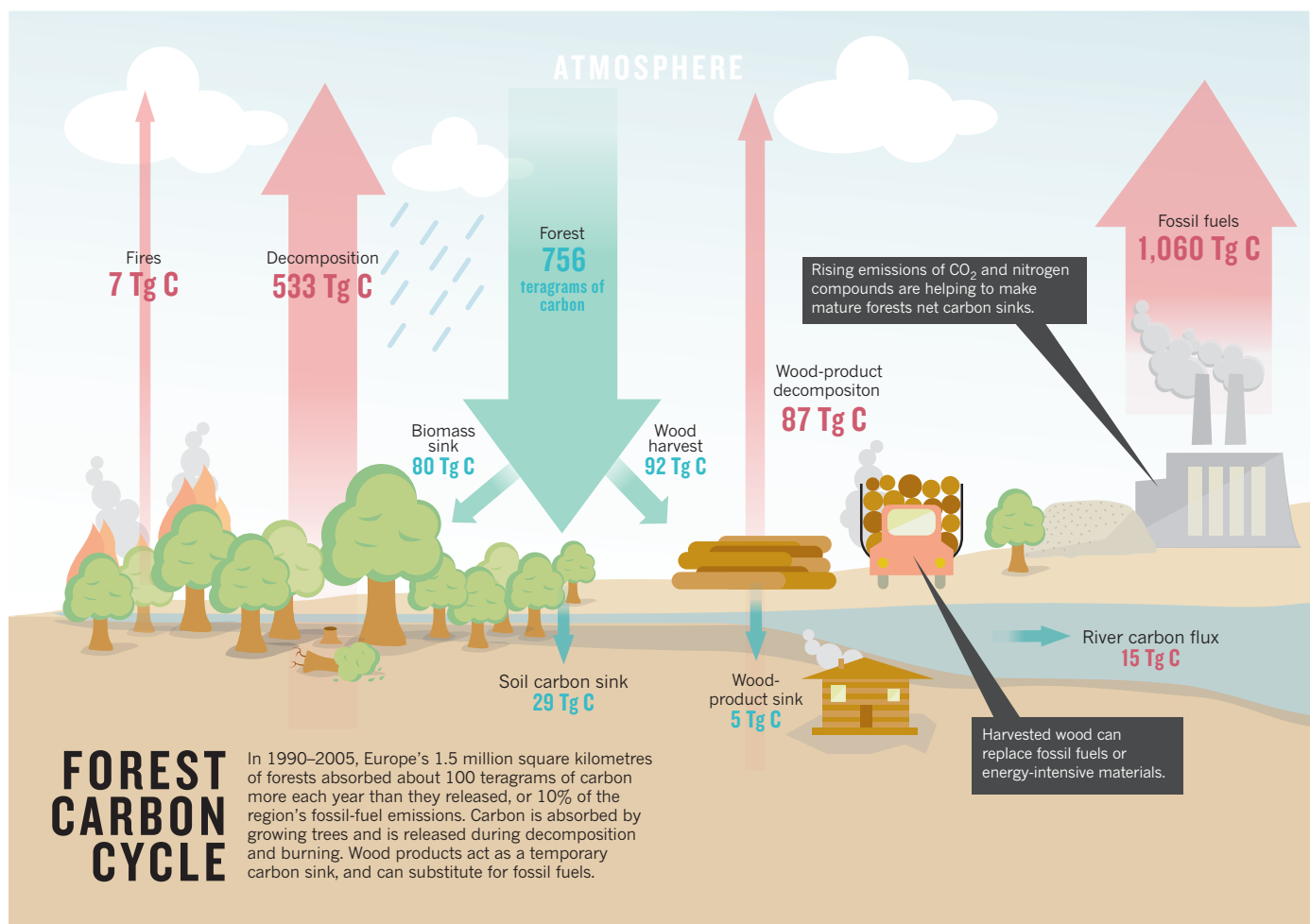
The best way to manage forests to store carbon and to mitigate climate change is hotly debated. Trees absorb carbon dioxide from the atmosphere, and wood can be a substitute for fossil fuels and carbon-intensive materials such as concrete and steel. In the past few decades, the world's forests have absorbed as much as 30% (2 petagrams of carbon per year; Pg C year^{-1}) of annual global anthropogenic CO_2 emissions¹

— about the same amount as the oceans. Two-thirds of forests are managed.

Much has been learned about the carbon cycle in forests, but there are still too many gaps in our knowledge. New observations have called long-accepted theories into question: the finding that unharvested forests, for example, are absorbing more carbon than they release², accounting for half the sink, is contrary to the tenet of ecology, known

as Odum's framework, that carbon flows in natural forests should be in equilibrium.

This carbon-sink behaviour of mature forests is attributed to large-scale environmental changes that violate the assumption of the steady conditions underlying Odum's framework: higher atmospheric CO_2 concentrations are accelerating tree growth worldwide and nitrogen emitted by industry, agriculture and fossil-fuel burning is ►



SOURCE: S. LUYSSAERT ET AL. GLOB. CHANGE BIOL. 16, 1429–1450 (2010)

► increasingly fertilizing managed forest soils in Europe, China and the eastern United States³.

To make good decisions about how to cultivate forests for climate-change mitigation, such as whether it is better to harvest or conserve trees, we must better understand the cause and future behaviour of this *in situ* carbon sink. Until more is known, we propose that forestry management should prioritize 'win-win' strategies — those that increase both forest stocks and timber harvest, through measures such as protecting trees from animals, or replacing dying or low-productivity forests (see 'Forest carbon cycle').

GLOBAL DRIVERS

So far, most discussions about the impact of climate change on forests have focused on the rising rate of locally devastating events such as forest fires, infestations, droughts and storms. The past decade saw 410 million cubic metres of wood felled in four major storms in Europe, a decade's worth of Amazonian carbon sequestration lost in severe droughts in 2005 and 2010, a record heat wave with forest fires affecting 23,000 square kilometres in Russia in 2010, and the bark beetle pandemic that has affected 130,000 square kilometres and

killed 435 million cubic metres of trees in British Columbia, Canada, since 2004.

Regionally, these events have had large socio-economic and ecological impacts. The windstorms that hit Europe in 1999, for example, destroyed one-third of the region's annual carbon sequestration⁴ by forests, and halved the price of timber in France and other affected countries. But such events are insignificant over decades and at the global scale. The Pacific El Niño and La Niña climate oscillations — primarily the tropical forest and peatland fires associated with them — are the only regional disturbances evident in globally averaged atmospheric CO₂ records⁵.

At the same time, atmospheric records and forest inventories show that forests have been taking up ever more CO₂ in the past 50 years¹. Experimental and modelling studies have established that currently, the global forest sink is driven mainly by changes in atmospheric CO₂ concentration and in nitrogen deposition. The competing explanation, that today's rapid tree growth is a recovery from ancient climate or

"Today's forest management is more of a gamble than a scientific debate."

harvesting losses, seems unlikely because it requires that regions around the globe, such as the Amazon and the Congo Basin, were simultaneously affected. Such a coincidence should have left traces of soot and enhanced CO₂ concentrations in air bubbles in glacial ice core samples, and these are not seen.

The size of the global forest carbon sink has increased alongside rising CO₂ levels (see 'Global land sink'). Industrial and agricultural activities and fossil-fuel burning emit nitrogen compounds that fertilize forests up to hundreds of kilometres away³. The rate of carbon storage in temperate forests in Europe and North America is well correlated with their exposure to this nitrogen deposition. Other changing climate factors, such as temperature and rainfall, and changing forest-management strategies, such as leaving trees to grow for longer before cutting them back, seem to be of secondary importance to the global carbon budget of forests, although they may be locally important^{6,7}.

A quantified understanding of how all these drivers shape the forest carbon sink is lacking. And predictions of how they will change during this century remain uncertain. Projections that CO₂ might overtake nitrogen deposition as the dominant cause of the sink in temperate zones, as

industrial emissions fall, are controversial. Physiological constraints, such as trees living shorter lives as their growth accelerates and soil fertility falling in mature woodlands, are often missing from models, as is the effect of temperature on tree growth.

As a result, the picture is confused. Different models disagree on whether the forest carbon balance in 2100 will be positive or negative, let alone its magnitude. Even where models make the same assumptions, such as no change in current emissions and no forest management, they have wildly different predictions. Models assuming that rising temperatures and CO₂ concentrations will increase photosynthesis, which absorbs CO₂ faster than respiration emits CO₂, suggest that the biosphere could absorb up to 10 PgC year⁻¹. This is five times today's terrestrial carbon sink and matches current CO₂ emissions from fossil-fuel burning and deforestation. If respiration dominates, others predict that the biosphere will become a carbon source of 6 PgC year⁻¹ (ref. 8), doubling the current CO₂ emissions and thus requiring emissions reductions far beyond what is being discussed.

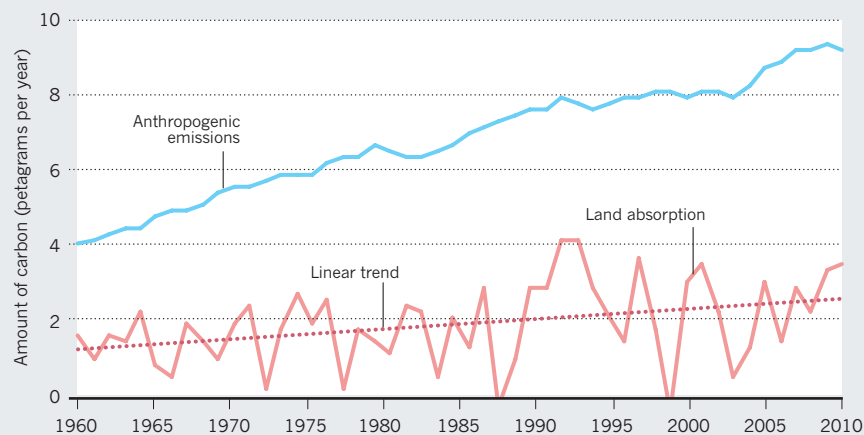
THREE WAYS OUT

The future trajectory of the carbon sink influences how forests should be managed for climate-change mitigation. If the world's forests remain net absorbers, conservation would be appealing. Preserved mature forests would absorb almost as much carbon as younger ones. Because decomposing harvest residues and roots add immediately to the CO₂ emissions, and it takes decades for increased use of wood products to compensate, avoiding harvest could generate extra climate benefits, at least in the short run.

Conversely, if mature forests become carbon sources, increased harvesting may be the best mitigation option. Harvesting would reduce losses from decomposition while promoting wood as a fossil-fuel substitute.

GLOBAL LAND SINK

The quantity of carbon absorbed by trees and other types of vegetation per hectare of land has risen in the past 50 years as anthropogenic carbon dioxide and nitrogen emissions have grown. This is despite the world's forest area falling by around 2% since 1990.



While a scientifically robust prediction of the persistence of the forest carbon sink is worked out, three safe strategies should be followed.

First, so that studies can be compared and uncertainties addressed, scientists should state their assumptions more clearly. The Earth system modelling community should define and report a set of internationally agreed performance quality criteria, based on an ability to reproduce trends in large data sets that include interactions relevant to understanding the effects of nitrogen and CO₂ on the forest sink. To provide these data sets, large-scale remote sensing needs to be combined with frequent monitoring of local test sites, such as the permanent plots of national forest inventories, with monitoring networks such as Fluxnet (which coordinates measurements of CO₂ exchanges), and with controlled experiments, such as CO₂-enrichment studies.

The forest-science community should make explicit the assumed behaviour of unmanaged forests that lies behind its assessment of forest mitigation strategies and its life-cycle analysis for wood products. Most of these studies⁹ assume that unmanaged forests are carbon neutral, which overestimates the climate benefits of harvest. Others¹⁰ assume that the *in situ* forest sink will be sustained forever, underestimating the climate benefits of harvest.

Second, the most carbon-efficient uses of wood should be encouraged. Harvesting more timber could be, especially if the forest sink starts shrinking, a good climate-change mitigation strategy, but to be effective it must be targeted to uses that will save the most tonnes of CO₂ per cubic metre harvested. In construction, for example, wood can substitute for steel or cement for studs or walls, and can then be recovered, recycled and burnt.

Third, forest-management techniques that increase both the amount of wood produced and the carbon stock retained in the forest should be prioritized. When not in conflict with other forest uses, replacing dying or low-productivity stands, protecting young sprouts from damage after harvest, planting tree mixes that are more resilient, and optimizing fertilizer use and tree growth by adding nitrogen-fixing species in afforestation projects, will contribute to climate-change mitigation no matter how the global carbon sink evolves.

Today's forest management is more of a gamble than a scientific debate. By following 'no-regret' strategies, we can buy time while we learn more. The future of the world's forest should not depend on tossing a coin. ■

Valentin Bellassen is at the Centre for Economics and Sociology in Rural Areas (CESAER), National Institute for Agronomic Research (INRA), Dijon, France. **Sebastiaan Luyssaert** is at the Laboratory for Climate Sciences and the Environment (LSCE), Atomic and Alternative Energy Commission (CEA), Gif-sur-Yvette, France. e-mail: valentin.bellassen@dijon.inra.fr

1. Pan, Y. et al. *Science* **333**, 988–993 (2011).
2. Luyssaert, S. et al. *Nature* **455**, 213–215 (2008).
3. Solberg, S. et al. *For. Ecol. Mgmt* **258**, 1735–1750 (2009).
4. Lindroth, A. et al. *Glob. Change Biol.* **15**, 346–355 (2009).
5. Reichstein, M. et al. *Nature* **500**, 287–295 (2013).
6. Bellassen, V. et al. *Glob. Change Biol.* **17**, 3274–3292 (2011).
7. Desai, A. R., Moorcroft, P. R., Bolstad, P. V. & Davis, K. J. *J. Geophys. Res.* **112**, G01017 (2007).
8. Friedlingstein, P. et al. *J. Clim.* **19**, 3337–3353 (2006).
9. Lippke, B. et al. *Carbon Mgmt* **2**, 303–333 (2011).
10. Hudiburg, T. W., Law, B. E., Wirth, C. & Luyssaert, S. *Nature Clim. Change* **1**, 419–423 (2011).



Harvesting reduces carbon emissions from decomposition.



GEOSTILLS/ALAMY

The Panamanian golden frog (*Atelopus zeteki*) is among thousands of species that have become critically endangered.

ENVIRONMENTAL SCIENCES

The long goodbye

A study of Earth's declining biodiversity deftly traces patterns and processes, finds **Bob Bloomfield**.

On page one of chapter one of Elizabeth Kolbert's *The Sixth Extinction*, the golden frogs of Panama are described as "taxicab yellow". Later, miles, acres and degrees Fahrenheit are mentioned. Coming as I do from a country that uses the International System of Units, and where cabs are blacker than the European starlings that Kolbert tells us have invaded her native New York, I began to wonder. But in the event, this is a truly global exploration of mass extinction.

Kolbert's 'hook' is a dramatic decline in biodiversity that has already begun and will inevitably accelerate over the coming decades. Her scope, however, is a wider synthesis of what we know about extinction. Kolbert intertwines her narrative with visits to the field sites and laboratories of leading scientists, which reveal the unfolding calamity. At a remote research station on Australia's Great Barrier Reef, for instance, she meets Ken Caldiera, who she says is credited with coining the term ocean acidification. And in the Manú National Park in Peru, she explores Miles Silman's altitudinal transects of a forest already responding to climate change. Each of the book's 13 chapters uses an emblematic species studied

by these scientists as a springboard for the discussion, although she has rightly drawn evidence from a much wider range.

Kolbert begins with the concept of extinction itself, which emerged in the eighteenth century. Until then, the belief that species were divinely created precluded the idea that they could be irretrievably lost. The realization that fossil deposits contained the remains of long-extinguished species arose amid a growing clash of ideas. Catastrophism, the brainchild of naturalist Georges Cuvier, allowed for revolutionary loss of fauna and flora, but did not provide an evolutionary explanation for their replacement.

Uniformitarianism, the concept of gradual, imperceptible geological change favoured by Charles Lyell, extended to Charles Darwin's explanation of evolution by means of natural selection — but sat uncomfortably with the idea of sudden, cataclysmic events.

Kolbert explores the

five mass extinctions that punctuate the story of life on Earth, deftly exploring the principal evidence of probable causes. She begins with the Cretaceous–Palaeogene boundary event 66 million years ago, and the controversial theory posited by theoretical physicist Luis Alvarez and his geologist son Walter that a meteor impact destroyed the ruling reptiles, making way for the rise of small mammals. For the Ordovician–Silurian mass extinction 444 million years ago, Kolbert looks at a possible cocktail of effects that triggered the loss of 85% of marine life. Was it the absorption of carbon dioxide by the first land plants that plunged Earth into an ice-house glaciation? The resulting ocean shrinkage, falling temperatures and increase in oxygen-altering marine chemistry would have devastated life.

Next, she examines the ecological consequences of human activity over the 40,000 years since the loss of Australia's megafauna. She points to the unprecedented scale and pace of our impacts from prehistory to today, including the appropriation of fossil fuels and the conversion of forests and other ecological systems, driven by demand for food and commodities. These will leave indelible traces — hence the



The Sixth Extinction — An Unnatural History
ELIZABETH KOLBERT
Henry Holt: 2014.

growing scientific consensus that we live in a new geological epoch, the Anthropocene. Kolbert drills into the mechanisms through which climate change, ocean acidity, ecological fragmentation and the movement of alien species are catastrophically acting on many populations. Those alien invasions are leading to what has been called a new Pangaea: just like on that supercontinent 300 million years ago, biodiversity is mixing — but this time it is causing a reduction in species.

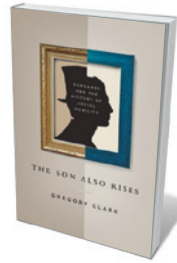
Kolbert's romp through these themes is necessarily selective, and some may take issue with aspects of her concise narrative. For example, Darwin did dismiss the sudden loss of species across boundaries between geological epochs as vast periods lost in an incomplete fossil record. But it seems somewhat disingenuous to implicate him in something that happened a century later: the resistance to the meteor-impact idea by George Gaylord Simpson and other palaeontologists of the hard-nosed uniformitarian tradition. Had Darwin been there, he would, I think, have based his response on evidence rather than dogma. As Kolbert herself summarizes, mass-extinction events occur on a grander scale than the inexorable chipping away of Darwin's 10,000 sharp wedges of nature. The Alvarezs pointed to a clear mechanism by which well-adapted organisms were swept aside and the pattern of life-forms upon which natural selection could act reset.

The Sixth Extinction lucidly introduces the context and process underlying the current doom-laden prognosis for life on Earth. In the penultimate chapter, Kolbert writes of the search for the bits of genetic code that endow humans with our unique restlessness, use of symbols, creativity and ability to work socially. We may be a “weedy species” but we have hugely changed our surroundings since we first appeared. The book's subtitle, *An Unnatural History*, holds only as long as we see ourselves outside of nature. The reality is that our future is not decoupled from that of the biosphere. The breaching of planetary boundaries — the potential tipping points in Earth systems such as climate — will have non-linear, unpredictable consequences for all species, including our own.

The five mass extinctions demonstrate that past success is no guarantee of survival: a catalogue of dominant species failed to bridge each of these “moments of panic”. Kolbert touches on, but stops short of probing, what we can do to make a difference for the future of all life on Earth. The possibilities for ecological adaptation and mitigation leave significant scope for a sequel. ■

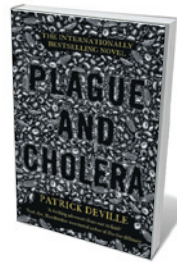
Bob Bloomfield is an independent specialist in public engagement with and understanding of science and the environment, based in London.
e-mail: bob_bloomfield@bobania.f9.co.uk

Books in brief



The Son Also Rises: Surnames and the History of Social Mobility Gregory Clark PRINCETON UNIVERSITY PRESS (2014)

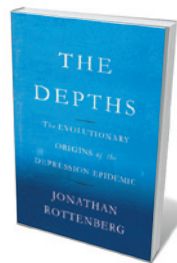
Bloody revolutions, policy upheavals and a deluge of social-science theories have been sparked by social inequality. So how porous is the class divide? Not very, reveals economic historian Gregory Clark in this audacious study based on tracking family names through history. Examining names in areas as far-flung in time and space as today's Sweden, Qing Dynasty China and medieval England, Clark shows how little social mobility has altered in 800 years. The solution to the status lottery, he argues, is for society to rectify the imbalance in rewards given to rich and poor.



Plague and Cholera

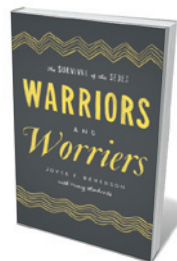
Patrick Deville (Translated by J. A. Underwood) LITTLE, BROWN (2014)

In 1894 Alexandre Yersin, a protégé of Louis Pasteur, discovered the bacillus behind a disease that had ravaged Europe for centuries: bubonic plague. It was dubbed *Yersinia pestis* in his honour. Patrick Deville's novel — a French best-seller now translated into English by J. A. Underwood — eloquently chronicles Yersin's eventful life. Starting with the scientist's flight out of Paris during the Second World War, the story swerves back in time to chronicle his explorations of Vietnam, the Philippines and “the new frontier of microbiology” in the tumult of the twentieth century.



The Depths: The Evolutionary Origins of the Depression Epidemic Jonathan Rottenberg BASIC BOOKS (2014)

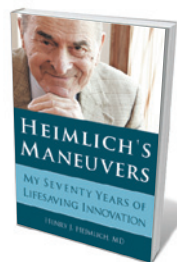
Depression, avers psychologist Jonathan Rottenberg, is an evolved trait: a way of stopping us in our tracks so that we can perceive the hurdles facing us. But in today's relatively safe environment, he argues, this useful adjunct of self-analysis can ramp up in severity — and strictly biological approaches and talking cures are not universally effective. Rottenberg brings clinical findings, experimental research, anecdotal evidence and personal experience of depression to his study of triggers, exacerbating factors, psychobiology and evidence for routes to recovery.



Warriors and Worriers: The Survival of the Sexes

Joyce F. Benenson with Henry Markovits OXFORD UNIVERSITY PRESS (2014)

In this provocative treatise, psychologist Joyce Benenson overturns the prevalent social-science theory that women are the more sociable sex and men more competitive. Benenson posits that the sexes exhibit the strongest differences in behaviours that support their long-term survival. The tendency among girls to discuss others is linked to the evolutionary need to sift out people who will help with childcare, she argues, but women often compete over men. Men, by contrast, cooperate in competing against other groups.



Heimlich's Maneuvers: My Seventy Years of Lifesaving Innovation Henry J. Heimlich PROMETHEUS BOOKS (2014)

Forty years ago, a thoracic surgeon first described a technique to halt choking by administering abdominal thrusts. Henry Heimlich's manoeuvre is now a globally recognized first-aid intervention. In this matter-of-fact memoir, Heimlich lays out a life spent crafting pragmatic health-care innovations. Along with inventions such as a ‘flutter valve’ to drain chest fluids, he touches on more controversial potential treatments, such as fighting HIV by inducing infection with malaria. [Barbara Kiser](#)



Rupert Till and his team capture the acoustic fingerprint of the Tito Bustillo cave in Spain.

Q&A Rupert Till

Acoustic archaeologist

Rupert Till at the University of Huddersfield, UK, studies the sonic properties of caves containing prehistoric paintings. As he addresses a conference in Malta on the archaeology of sound, he talks about the hum of Stonehenge, acoustic fingerprinting and simulating primeval concerts in the dark.



How did you come to study cave acoustics?

The anthropologist Igor Reznikoff believes that cave paintings were sited on the basis of the acoustic properties of those chambers, and

says he can locate the paintings in complete darkness by using his voice to gauge the resonance of the spaces. Last year, I travelled with a group of archaeologists and musicians to explore this hypothesis in five caves in northern Spain. In both painted and unpainted chambers, we used a laptop and loudspeaker to sweep a sine wave tone through all audio frequencies, recording the results to capture the acoustic fingerprint of each space. This can give hints about what might have occurred there: how intelligible speech would have been; whether the chamber would have been suited to music; and whether it might have been ritually important due to unusual sounds or acoustic effects.

What patterns did you find?

Our acoustic testing revealed a change in the location of paintings over the millennia. The oldest paintings, from up to 40,000 years ago — some as simple as dots or handprints — tend to be in small, intimate places where there is less reverberation. Perhaps 15,000 to 20,000 years later we get paintings of animals like deer and bison, sometimes overlaid on top of each other, starting to appear in more echoey spaces that are large enough for groups of people to have gathered for rituals.

What kind of music do you think people made in these caves?

We may have to think about sound-making, rather than music with performers and audiences. There is generally little background noise, so even the sound of footsteps might have seemed loud. If there were rituals, everyone might have participated by clapping, stomping, banging or singing.

Have you experimented with playing ancient instruments *in situ*?

Some of the earliest musical instruments have been found in caves. We brought simple tools like bone scrapers, seashells and river stones.

Archaeoacoustics: The Archaeology of Sound

19–22 February 2014, Corinthia Palace Hotel, Malta.

Bullroarers — leaf-shaped pieces of stone or wood that emit powerful low frequencies as you whirl them on a string — have been found in a nearby cave, so we brought those too. At one point, someone played a reconstruction of a 40,000-year-old vulture-bone flute in complete darkness. The music seemed to bring the environment to life.

Might natural sounds have formed part of the ancient soundscape?

Mostly the caves are incredibly quiet. Near some dramatic paintings there are big holes in the ground that open onto underground rivers whose sound rushes up from the depths. Some stalagmites and stalactites ring like a xylophone when struck, and were marked with paint in prehistory. In a way, a cave plays itself: you'll hear a little ping as water drips from the ceiling onto a stalagmite.

Can the public access prehistoric music?

We created a film aiming to reconstruct what it might have looked, sounded and felt like to be in the caves in prehistory, which will be on the Songs of the Caves project website (<http://songsofthecaves.wordpress.com>). In 2015–16, as part of the European Music Archaeology Project, we will mount a travelling exhibition about ancient instruments, including simulations of spaces such as caves and temples, and recordings of instruments being played with accurate acoustics. I'd love to document the acoustics in monuments such as the Taj Mahal and the Parthenon.

What did you discover about Stonehenge?

I worked with acoustician Bruno Fazenda to reconstruct how Stonehenge would have sounded when it was intact. Using digital modelling, we found that the ring of stone slabs had two modes of resonance that line up near the lowest F sharp on a grand piano. If the wind blew hard, it could produce a powerful hum at that frequency. Thomas Hardy wrote about the phenomenon in his novel *Tess of the d'Urbervilles*: "The wind, playing upon the edifice, produced a booming tune, like the note of some gigantic one-stringed harp." It's hard to hear now because of traffic.

What will remain of today's music?

It's hard to say. When we study ancient music, the instruments we find may just be the ones made of materials that lasted. But it may be that most music was made by clapping and singing, or by striking clay pots or pieces of wood, which leave no archaeological trace after a certain period. I'm sure that in thousands of years people will have a distorted image of what music means to us today. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Biodiversity panel should play by rules

The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) is adopting its initial work programme for 2014–18. This is a good time to assess whether the “rules of engagement” outlined for its success (E. Turnhout *et al. Nature* **488**, 454–455; 2012) have been put into practice. We find that there are still some serious shortfalls that, unless addressed promptly, could undermine the transparency and credibility of the output from the IPBES.

Some of the nine rules (R1–R9; see go.nature.com/guc1gk) are now being followed, at least partly: reporting on progress is to be staggered (R8); the call for a decentralized approach that is sensitive to local knowledge, needs and conditions (R1) is now reflected in several deliverables; and most mandated functions are being addressed simultaneously (R2). Cautious optimism is also justified for R3 and R5: expert groups for discussing terms and methodologies are to be established, and non-scientific knowledge systems will be included.

However, non-elite actors are not yet properly involved, and decisions on stakeholder engagement and communication strategies have been postponed, which will affect R3 and R5; this also applies to the inclusion of “trusted civic organizations and networks” (R7).

Furthermore, stakeholders are only marginally involved in nomination procedures for the Multidisciplinary Expert Panel (MEP) and for experts who will work on IPBES deliverables, and are not involved at all in the selection process (R4). Of the MEP members selected, the mix is unbalanced (22 natural scientists, 2 economists and 1 social scientist, and women are under-represented). It is also unclear whether minority dissenting views will be incorporated into the IPBES

process (R6). Re-evaluation and improvement of IPBES processes (R9) will be hard because IPBES members have declared that they are not prepared to renegotiate after issues have been formally agreed.

Lars Opgenoorth, Stefan Hotes
University of Marburg, Germany.
opgenoorh@uni-marburg.de
Harold Mooney *Stanford University, California, USA.*

Sexual violence rife on US campuses

Gun violence may be relatively rare on university campuses in the United States (see *Nature* **505**, 150–152; 2014), but other violent incidents are rampant.

Acts of sexual violence, for example, are committed against an estimated 20–25% of females attending a US college or university (see C. J. Vladutiu *et al. Trauma Violence Abus.* **12**, 67–86; 2011). According to the US Census Bureau, roughly 11,658,000 women were enrolled in higher-education institutions in 2009. Assuming 20% victimization, as many as 2,331,600 of those could have been sexually assaulted on campus.

You missed an opportunity to expand on a horrifying problem that is all too often under-reported.

Stephen Shaffer *Endicott, New York, USA.*
smshaffe@gmail.com

Scientists need leadership training

Career scientists need to be able to manage projects, students, postdocs, staff and, ultimately, a scientific team. We suggest that junior scientists should routinely be trained in such leadership skills to maximize the productivity of their future research groups (see also J. C. Seeliger *Nature* **483**, 511; 2012).

Unfortunately, many scientists fail to realize that they need

training, and there are only a few leadership programmes on offer. One is the three-day ‘Leadership in BioScience’ workshop for young investigators held annually by the Cold Spring Harbor Laboratory in New York (see go.nature.com/ocx5yp). Through lectures, role-playing exercises, case studies and discussions, participants learn and practise skills such as setting goals, delivering feedback, running successful meetings and managing conflict or difficult situations that can arise in the laboratory. Training is consolidated with detailed constructive feedback from instructors and other participants.

More such programmes should be developed to help junior faculty members in academic institutions worldwide.

Marina Kvaskoff *Brigham and Women’s Hospital, Boston, Massachusetts, USA.*
marina.kvaskoff@channing.harvard.edu

Stephanie D. McKay *University of Vermont, Burlington, USA.*

Cuts threaten future of R&D in Portugal

Drastic cuts in public funding for science and technology in Portugal are being applied as part of the government’s sweeping austerity measures, without thought for a national long-term science strategy. The effects are particularly hard on young scientists.

The number of PhD and postdoctoral fellowships awarded by the country’s principal funder, the Science and Technology Foundation (FCT), fell by 40% from 2012 to 2013, and the 2014 budget for fellowships fell by 16.5% compared with last year.

Coupled with reductions in funding for research centres and research and development (R&D) projects, these cuts have serious long-term implications for the country’s chance of

developing a knowledge-based economy.

André Levy *ISPA, Lisbon, Portugal.*
andre_levy@ispa.pt

Online repository for lab notebooks

The US Office of Science and Technology Policy (OSTP) ruled last year that all federally funded research articles and data should be open access, but laboratory notebooks were not included (see go.nature.com/fijt2f). I urge researchers to sign a petition (see <http://wh.gov/l5gv0>) to extend this mandate to cover a searchable online repository of key lab notebooks, such as those used in formulating US patent applications.

The legal value of lab notebooks is complicated, which may be one reason for their omission from the OSTP mandate. However, freely accessible notebooks are useful as evidence in patent litigation and can help to resolve irreproducibility issues that might otherwise have economic and medical implications for products and processes. They also act as an archive of raw data for the scientific and medical record.

As an agency that is fully funded by user fees and one that aims to validate patent findings, perhaps the US Patent and Trademark Office would consider financing the initial purchase of server space and set-up costs for a notebook repository.

Shannon Bohle *Lima, Ohio, USA.*
sb838@cantab.net

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at go.nature.com/cmchno. Alternatively, readers may comment online: www.nature.com/nature.

Uncertainties in transpiration estimates

ARISING FROM S. Jasechko *et al.* *Nature* **496**, 347–350 (2013)

How best to assess the respective importance of plant transpiration over evaporation from open waters, soils and short-term storage such as tree canopies and understories (interception) has long been debated. On the basis of data from lake catchments, Jasechko *et al.*¹ conclude that transpiration accounts for 80–90% of total land evaporation globally (Fig. 1a). However, another choice of input data, together with more conservative accounting of the related uncertainties, reduces and widens the transpiration ratio estimation to 35–80%. Hence, climate models do not necessarily conflict with observations, but more measurements on the catchment scale are needed to reduce the uncertainty range. There is a Reply to this Brief Communications Arising by Jasechko, S. *et al.* *Nature* **506**, <http://dx.doi.org/10.1038/nature12926> (2014).

The selected values for runoff Q and interception xP in equation (4) of ref. 1 are low, resulting in high transpiration ratios. After consulting multiple state-of-the-art studies and using the mean and standard deviation (σ) of these estimates, we estimate Q at $39,600 \pm 5,100 \text{ km}^3$ per year (refs 2–4) and xP at $20,100 \pm 9,800 \text{ km}^3$ per year (refs 5–7), even without including understory interception⁸. The use of these more realistic input data lowers the transpiration ratio to 50–80% (Fig. 1b).

To estimate the isotopic composition of transpired moisture, δ_T is split between shallow and deep waters on the basis of normalized difference vegetation indices and precipitation. Although we agree that δ_T is difficult to estimate, we do not see a compelling physical reason for using this particular method to estimate deep δ_T , because the assumption that deeper water reflects the isotopic composition of rainfall has not been justified.

Rain water percolating to greater depths has to pass the unsaturated zone, where the isotopic composition of temporally stored water depends on evaporation (and thus fractionation) from interception storage (for example, forest floor⁸) and the topsoil⁹. Consequently, percolation through the unsaturated zone allows the rain water to mix with old, fractionated

water, which results in deeper water having a different isotopic composition from rain water. Where preferential percolation through macropores is dominant, the assumption may be justifiable because water may bypass the soil matrix without significant mixing^{9,10}. However, preferential percolation is not everywhere and always dominant.

Soil and open-water evaporation are isotopically grouped into one term and the estimate of the isotopic composition of evaporate is based on a very sensitive laboratory-derived evaporation model. The ‘constant’ C_K (equation (8) in ref. 1), however, depends on local climatic and physiographic conditions¹¹. The applicability of this laboratory-derived model to global estimates has not been shown, nor has its sensitivity to the use of grid- and time-averaged values for the four inputs (T_A , h_A , δ_A and T_L) yet been assessed. The sensitivity analysis was performed with arbitrarily fixed values of σ for T_A , h_A , T_L and δ_A .

Besides the estimates of δ_T and δ_E (and thus d_T and d_E), we also consider the uncertainty ranges of ref. 1 far too optimistic. Performing an error propagation on the deuterium excess value, we find $\sigma^2(d) = \sigma^2(\delta^2\text{H}) + 8^2\sigma^2(\delta^{18}\text{O})$. For $\sigma(\delta^2\text{H}_T) = 14\text{‰}$ and $\sigma(\delta^{18}\text{O}_T) = 1.7\text{‰}$ (Supplementary Table 4 in ref. 1), we find $\sigma(d_T) = 19\text{‰}$. Hence, it is not clear how the authors estimated $\sigma(d_T) = 3\text{‰}$. Accounting for all errors in equation (8) will result in a much higher estimate for $\sigma(d_E)$ than 30‰. Nonetheless, neglecting the propagation of errors and as an illustration just doubling the uncertainty in d_E , we find that the transpiration ratio decreases to 35–80% (Fig. 1c).

The fact that the transpiration ratio of ref. 1 aligns with a gross primary production estimate via the water-use efficiency (WUE) is not conclusive. First, the uncertainties in global WUE are high, owing to non-representative point measurements and the neglect of the underlying uncertainties in the WUE-regression relations (Supplementary Fig. 5 in ref. 1). Moreover, WUE varies in time, for example, owing to the CO_2 -fertilization effect¹². Second, gross primary production estimates are highly uncertain because gross primary production models often fail at the regional and global scale, owing to spatial and temporal heterogeneity, nonlinearity and site-specific data requirements¹³. Moreover, validating these models is problematic, because there are no direct measures at scales larger than the leaf level¹⁴.

Although isotopes are useful for partitioning evaporation at the small scale¹⁵, results at the larger scale are highly sensitive to the input data and their related uncertainties. Future improvements in observation techniques will enhance the application of isotopes to the partitioning of evaporative fluxes.

A. M. J. Coenders-Gerrits¹, R. J. van der Ent¹, T. A. Bogaard¹, L. Wang-Erlandsson^{1,2}, M. Hrachowitz¹ & H. H. G. Savenije¹

¹Water Resources Section, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands.

email: a.m.j.coenders@tudelft.nl

²Stockholm Resilience Centre, Kräftriket 2, SE 10691 Stockholm, Sweden.

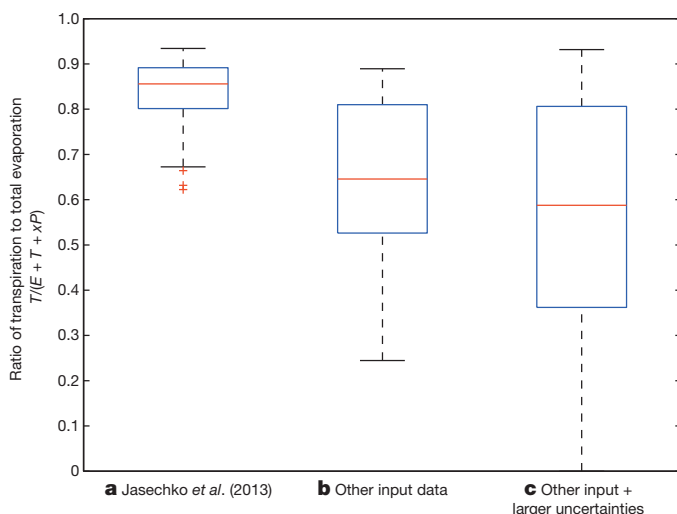


Figure 1 | Ratio of transpiration to total evaporation. a–c, Box plots are calculated using a simplified Monte Carlo simulation of equation (4)¹ with data from Jasechko *et al.*¹ (a), and with the same data as in a but with $Q = 39,600 \pm 5,100 \text{ km}^3$ per year and $xP = 20,100 \pm 9,800 \text{ km}^3$ per year (b), and with the same data as in b but with $d_E = 75 \pm 60\text{‰}$ (c). The blue box indicates the 25th and 75th percentiles with the median in red. The error bars indicate the minimum and maximum values. The red crosses indicate outliers (3/2 times the central box).

1. Jasechko, S. *et al.* Terrestrial water fluxes dominated by transpiration. *Nature* **496**, 347–350 (2013).
2. Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068–1072 (2006).
3. Dai, A. & Trenberth, K. E. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *J. Hydrometeorol.* **3**, 660–687 (2002).
4. Syed, T. H., Famiglietti, J. S., Chambers, D. P., Willis, J. K. & Hilburn, K. Satellite-based global-ocean mass balance estimates of interannual variability and emerging trends in continental freshwater discharge. *Proc. Natl Acad. Sci. USA* **107**, 17916–17921 (2010).

- Wang, D., Wang, G. & Anagnostou, E. N. Evaluation of canopy interception schemes in land surface models. *J. Hydrol.* **347**, 308–318 (2007).
- Dirmeyer, P. A. *et al.* GSWP-2: multimodel analysis and implications for our perception of the land surface. *Bull. Am. Meteorol. Soc.* **87**, 1381–1397 (2006).
- Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M. & Dolman, A. J. Global canopy interception from satellite observations. *J. Geophys. Res.* **115**, 27 (2010).
- Gerrits, A. M. J., Pfister, L. & Savenije, H. H. G. Spatial and temporal variability of canopy and forest floor interception in a beech forest. *Hydrol. Processes* **24**, 3011–3025 (2010).
- Brooks, J. R., Barnard, H. R., Coulombe, R. & McDonnell, J. J. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geosci.* **3**, 100–104 (2010).
- Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D. & Soulsby, C. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? *Hydrol. Earth Syst. Sci.* **17**, 533–564 (2013).
- Kendall, C. & McDonnell, J. J. *Isotope Tracers in Catchment Hydrology* (Elsevier, 1998).
- Keenan, T. F. *et al.* Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499**, 324–327 (2013).
- Yuan, W. *et al.* Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sens. Environ.* **114**, 1416–1431 (2010).
- Welp, L. R. *et al.* Interannual variability in the oxygen isotopes of atmospheric CO₂ driven by El Niño. *Nature* **477**, 579–582 (2011).
- Zhang, S., Wen, X., Wang, J., Yu, G. & Sun, X. The use of stable isotopes to partition evapotranspiration fluxes into evaporation and transpiration. *Acta Ecol. Sin.* **30**, 201–209 (2010).

Author Contributions All authors contributed equally, and all authors contributed to the data analysis.

Competing Financial Interests Declared none.

doi:10.1038/nature12925

Jasechko *et al.* reply

REPLYING TO A. M. J. Coenders-Gerrits *et al.* *Nature* **506**, <http://dx.doi.org/10.1038/nature12925> (2014)

In their Comment, Coenders-Gerrits *et al.*¹ suggest that our conclusion that transpiration dominates the terrestrial water cycle² is biased by unrepresentative input data and optimistic uncertainty ranges related to runoff, interception and the isotopic compositions of transpired and evaporated moisture. We clearly presented the uncertainties applied in our Monte-Carlo sensitivity analysis, we reported percentile ranges of results rather than standard deviations to best communicate the non-linear nature of the isotopic evaporation model, and we highlighted that the uncertainty in our calculation remains large, particularly in humid catchments (for example, figure 2 in our paper²).

A global runoff flux of 37,288 km³ per year was obtained from a previous compilation of discharge measurements³, and has been independently reproduced within 1,300 km³ per year (36,055 km³ per year) using satellite measurements⁴. The larger uncertainty reported by Coenders-Gerrits *et al.*¹ ($\pm 5,100$) is obtained using a higher runoff value from a schematic in a review⁵, which also included direct groundwater discharge to oceans, making this number 10% higher than it should be. Because the uncertainty estimate recommended by Coenders-Gerrits *et al.*¹ also includes water fluxes not included in our runoff parameter, we stand by the original values used, which seem to us to be the best approximation of this uncertainty^{3,4}.

Global interception, derived from satellite measurements^{6,7} (7,500 km³ per year, as used in our paper²), global climate model outputs⁸ (11,900 km³ per year, shown in figure 6b of ref. 4) and stand-level measurements⁹ (median value of 17.9% of precipitation, or 19,700 km³ per year), all fall into a range from 7,500 km³ per year to 19,700 km³ per year. We note that the high end-member applied by ref. 1 is derived from stand-level measurements and was not intended to be used to estimate a global interception flux. Its use in this manner¹ introduces bias towards regions where interception is expected to be important, because interception measurements are not often reported in areas free of canopy cover.

The Comment¹ also suggests that the uncertainties we applied to the deuterium excess ($d = \delta^2\text{H} - 8\delta^{18}\text{O}$) of transpired moisture (d_T) and evaporate (d_E) were understated². First, Coenders-Gerrits *et al.*¹ propose a deuterium excess uncertainty of $\pm 19\%$ for transpired moisture, quoting the individual uncertainties in our paper in the $\delta^{18}\text{O}$ and $\delta^2\text{H}$ models. This uncertainty is too large because they do not recognize that $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values covary¹⁰, resulting in a well-constrained range of deuterium excess values for groundwater tapped by plant roots. The uncertainty we applied² for d_T ($\pm 3\%$) is validated by a database of stable oxygen and hydrogen isotope values for shallow and deep groundwater across the continental USA (the Water Quality Portal, <http://www.waterqualitydata.us>). Samples from depths of <4.6 m—the global average

rooting depth¹¹—show a deuterium excess of $7.8 \pm 4.1\%$ ($n = 1,021$; 25th to 75th percentile range of data set), found to be nearly identical in deuterium excess to that of samples collected from depths greater than 4.6 m: $d = 7.9 \pm 3.1\%$ ($n = 24,309$). These ranges are consistent with our estimate² of d_T of $8 \pm 3\%$. Coenders-Gerrits *et al.*¹ suggest that the deuterium excess value ascribed to evaporating moisture should be higher than $\pm 30\%$. Our uncertainty of $\pm 30\%$ was derived from the range of deuterium excess calculations for 73 lakes used in our study. We used percentile ranges so that the high sensitivity (and associated high uncertainty) would be carried through to our final, global scale calculation of transpiration and evaporation fluxes.

Finally, Coenders-Gerrits *et al.*¹ suggest that we neglect uncertainty in water-use efficiency (WUE), despite our use of the standard error of regression that resulted in a global WUE of 3.4 ± 0.9 mmol CO₂ per mol H₂O (ref. 2). CO₂ fertilization does indeed affect WUE¹², as we acknowledged², and continued research into WUE will allow us to monitor and map these changes as atmospheric CO₂ increases. We strongly disagree with Coenders-Gerrits *et al.*¹ that gross primary production models “fail at the regional and local scale,” given the success of the FLUXNET project in constraining this vital component of the carbon cycle (123 ± 8 Gt of C per year reported¹³). Finally, WUE has in fact been measured at scales that are larger than the leaf level (see Supplementary Information refs 105, 114, 116 and 118 in ref. 2).

We do not agree with the magnitude of the increase in uncertainty in ref. 1, nor with the unidirectional changes to input parameters that produce lower transpiration fluxes. However, we agree with the final point of the Comment¹: that further method development and the use of stable O and H isotopes in hydrology can help to reduce uncertainty in terrestrial evaporation and transpiration.

Scott Jasechko¹, Zachary D. Sharp¹, John J. Gibson^{2,3}, S. Jean Birks^{2,4}, Yi Yi^{2,3} & Peter J. Fawcett¹

¹Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA.

email: jasechko@unm.edu

²Alberta Innovates—Technology Futures, Vancouver Island Technology Park, Victoria, British Columbia V8Z 7X8, Canada.

³Department of Geography, University of Victoria, Victoria, British Columbia V8W 3R4, Canada.

⁴Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

- Coenders-Gerrits, A. M. G. *et al.* Uncertainties in transpiration estimates. *Nature* **506**, <http://dx.doi.org/10.1038/nature12925> (2014).

BRIEF COMMUNICATIONS ARISING

2. Jasechko, S. *et al.* Terrestrial water fluxes dominated by transpiration. *Nature* **496**, 347–350 (2013).
3. Dai, A. & Trenberth, K. E. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *J. Hydrometeorol.* **3**, 660–687 (2002).
4. Syed, T. H., Famiglietti, J. S., Chambers, D. P., Willis, J. K. & Hilburn, K. Satellite-based global-ocean mass balance estimates of interannual variability and emerging trends in continental freshwater discharge. *Proc. Natl Acad. Sci. USA* **107**, 17916–17921 (2010).
5. Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068–1072 (2006).
6. Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M. & Dolman, A. J. Global canopy interception from satellite observations. *J. Geophys. Res.* **115**, D16122 (2010).
7. Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H. & Dolman, A. J. Magnitude and variability of land evaporation and its components at the global scale. *Hydrol. Earth Syst. Sci.* **15**, 967–981 (2011).
8. Dirmeyer, P. A. *et al.* GSWP-2: multimodel analysis and implications for our perception of the land surface. *Bull. Am. Meteorol. Soc.* **87**, 1381–1397 (2006).
9. Wang, D., Wang, G. & Anagnostou, E. N. Evaluation of canopy interception schemes in land surface models. *J. Hydrol. (Amst.)* **347**, 308–318 (2007).
10. Friedman, I. Deuterium content of natural waters and other substances. *Geochim. Cosmochim. Acta* **4**, 89–103 (1953).
11. Canadell, J. *et al.* Maximum rooting depth of vegetation types at the global scale. *Oecologia* **108**, 583–595 (1996).
12. Betts, R. A. *et al.* Projected increase in continental runoff due to plant responses to increasing carbon dioxide. *Nature* **448**, 1037–1041 (2007).
13. Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329**, 834–838 (2010).

doi:10.1038/nature12926

Genetic roots of the first Americans

The whole-genome sequence of a human associated with the earliest widespread culture in North America confirms the Asian ancestry of the Clovis people and their relatedness to present-day Native Americans. [SEE LETTER P.225](#)

JENNIFER A. RAFF & DEBORAH A. BOLNICK

Advances in genetics and archaeology over recent decades have transformed our understanding of the ancestry and migration history of indigenous Americans^{1–5}. However, limited genetic information from the earliest inhabitants of the Americas (those living more than 5,000 years before present) has made it difficult to test many hypotheses about the settlement and early history of this continent. On page 225 of this issue, Rasmussen *et al.*⁶ describe the complete genome sequence of an infant boy, Anzick-1, who was interred between 12,707 and 12,556 years ago at the Anzick site in Montana. Anzick-1 is the only known human burial associated with the Clovis culture — the oldest widespread archaeological complex in North America⁷, which is dated to 13,000 to 12,600 years ago. His genome is the first from an ancient American individual to be fully sequenced.

Anzick-1's mitochondrial DNA (maternally inherited extra-nuclear DNA located in cell mitochondria) belongs to the D4h3a lineage, a pan-American lineage thought to be one of the 'founder' lineages carried by the first people to enter the Americas. It is rare among present-day Native Americans (found in 1.4% of individuals) but more common in the oldest inhabitants of the Americas^{8,9}. Anzick-1's paternally inherited Y-chromosome lineage (Q-L54*(xM3)) is common in Native Americans¹. Comparisons of his nuclear genome with genetic data from Eurasian and Native American populations show that he is most closely related to Native American populations and more genetically similar to Siberians than to other Eurasians.

These genetic affinities between Anzick-1 and other Native Americans are not a surprise, because there is ample evidence that contemporary Native Americans are descended from the ancient inhabitants of the Americas^{1–3,10}. However, debate has persisted about this relationship because some of the earliest American skeletons exhibit distinctive craniofacial features¹¹ and because of continued attention to the Solutrean hypothesis, which proposes that the Clovis complex was derived from a transatlantic migration of Europeans who used Solutrean tools between 21,000 and

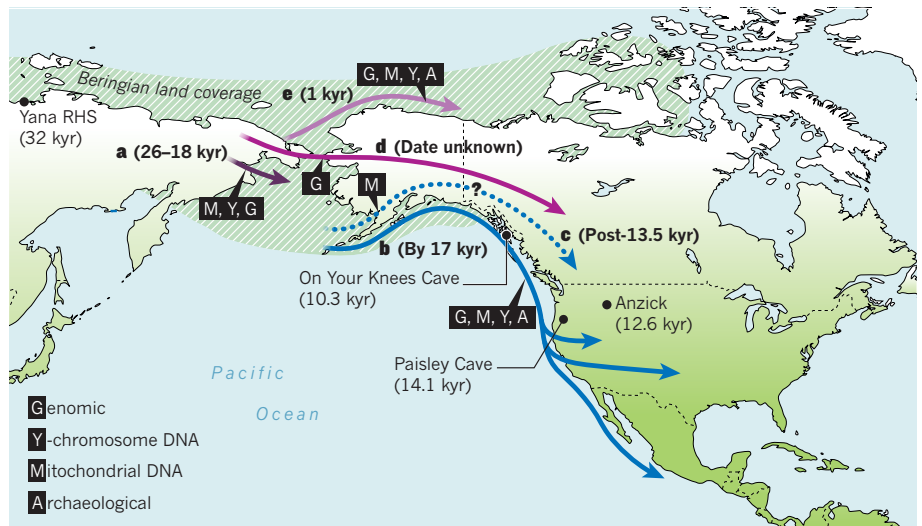


Figure 1 | Populating a continent. A hypothetical scenario for the peopling of the Americas, showing possible migration events (a–e) as described in the main text, coloured according to putative region of origin (Beringia, blue; Siberia, shades of purple). Also shown are some key archaeological sites in North America and Siberia, and the type of evidence (genomic, mitochondrial DNA, Y-chromosome DNA and archaeological) currently supporting each hypothesized migration. Shading depicts the extent of Beringia during the Last Glacial Maximum. (kyr, thousand years ago.)

17,000 years ago¹². The fragmentary nature of Anzick-1's remains means that they cannot be used to address questions about whether ancient Americans with distinctive craniofacial morphology were related to modern Native Americans. However, this genome contributes to a large body of genetic and archaeological research to lay the Solutrean hypothesis to rest once and for all. It shows that Anzick-1 indisputably had Asian ancestors and is related to contemporary Native Americans. Unless proponents offer evidence of direct European ancestry in other securely dated ancient American genomes, the Solutrean hypothesis can no longer be treated as a credible alternative for Clovis (or Native American) origins. It is time to move on to more interesting questions.

Rasmussen and colleagues' analysis of the Anzick-1 genome highlights some such questions. They find support for the proposal that Native Americans are descended from three 'streams' of gene flow from Asia, with some North Americans tracing their ancestry to multiple independent migrations from Siberia⁵. Intriguingly, their study reveals additional genetic substructure in North America — some

Native North American populations are genetically differentiated from Central and South Americans and from Anzick-1, indicating ancient divergences between Native American populations. The authors rule out additional migration from Eurasia as the cause of these patterns, but more North American genomes are needed to elucidate the evolutionary history that produced this structure.

The Anzick-1 genome, together with other genetic and archaeological evidence^{1–10}, helps to build a possible scenario for the peopling of the Americas. We know that humans had moved into northwestern Beringia from Siberia by 32,000 years ago¹³. Sometime between 26,000 and 18,000 years ago, these people expanded into eastern Beringia (Fig. 1a) and evolved genetic characteristics unique to Native American populations¹⁴. By 17,000 years ago, deglaciation of the Pacific coastline allowed a southward coastal migration (Fig. 1b), bringing the first Americans to Monte Verde in Chile by 14,600 years ago^{15,16}. Because Anzick-1's mitochondrial lineage is primarily found along the Pacific coast, this migration may have included his ancestors,

the ancestors of Central and South Americans, and those of other D4h3a-bearing populations (although we note that Rasmussen *et al.* may not agree with this interpretation). Descendants of these migrants who moved eastward, south of the Cordilleran and Laurentide ice sheets, could have developed the Clovis culture³. Some mitochondrial DNA studies also suggest a second wave of migration through the interior ice-free corridor that opened after 13,500 years ago¹⁷ (Fig. 1c), but this hypothesis is not yet supported by other evidence.

Genome-wide (and possibly mitochondrial) evidence^{5,7} shows that a second group of Siberians subsequently moved into North America (Fig. 1d), contributing to the ancestry of modern Chipewyans and perhaps other North American groups. Another expansion from Siberia brought Neo-Eskimo ancestors to North America (Fig. 1e) sometime before their expansion eastward around 800–1,000 years ago^{1,10,18}. Thousands of years of subsequent gene flow and other evolutionary forces, as well as region-specific cultural histories, contributed to the population structure of indigenous Americans.

The Anzick-1 study — similar to earlier studies of the 4,000-year-old Saqqaq genome from Greenland¹⁸, the 24,000-year-old Mal'ta genome from Siberia¹⁹ and the 7,000-year-old Mesolithic genome from Spain²⁰ — demonstrates that ancient genomes can yield insights into prehistoric population histories. However, the conclusions that can be drawn from a single genome are limited. Genome sequencing is not yet cost-effective for population-scale sampling of ancient remains, and many ancient samples do not contain enough DNA for complete genome sequencing. Studies of targeted DNA regions are therefore still needed. Furthermore, there are limitations on what we can learn from genetics alone, necessitating models that account for all available palaeoclimatological, genetic and archaeological data.

The Anzick-1 genome confirms many predictions from archaeological and genetic studies, and moves us closer to a nuanced and comprehensive model for the population history of the Americas. However, this genome, similar to the Mal'ta¹⁹ and European Mesolithic²⁰ genomes, shows that patterns of genetic diversity have changed dramatically over time. Some genetic variants that are rare today were more common in the past, and variants that are ubiquitous today may have increased in frequency only in the past few thousand years. Moreover, the current distribution of genetic lineages may not reflect the geographic areas in which they were found thousands of years ago — some lineages clearly had a different or wider range in the past^{6,19}. Perhaps the most important lesson from palaeogenomic studies, then, is that we must be cautious about assuming that present-day genetic patterns are representative of the past. Analyses of ancient human genomes are making it clear that

thousands of years of evolutionary processes have drastically reshaped the genetic landscape within any given geographic region. ■

Jennifer A. Raff and Deborah A. Bolnick
are in the Department of Anthropology,
University of Texas at Austin, Austin, Texas
78712-1723, USA.
e-mails: jenniferraff@utexas.edu;
deborah.bolnick@austin.utexas.edu

1. O'Rourke, D. H. & Raff, J. A. *Curr. Biol.* **20**, R202–R207 (2010).
2. Kemp, B. M. & Schurr, T. G. in *Human Variation in the Americas* (ed. Auerbach, B. M.) 12–50 (Southern Illinois Univ., 2010).
3. Goebel, T., Waters, M. R. & O'Rourke, D. H. *Science* **319**, 1497–1502 (2008).
4. Gilbert, M. T. P. *et al. Science* **320**, 786–789 (2008).
5. Reich, D. *et al. Nature* **488**, 370–374 (2012).

6. Rasmussen, M. *et al. Nature* **506**, 225–229 (2014).
7. Waters, M. R. & Stafford, T. W. Jr *Science* **315**, 1122–1126 (2007).
8. Kemp, B. M. *et al. Am. J. Phys. Anthropol.* **132**, 605–621 (2007).
9. Bolnick, D. *et al. Am. J. Phys. Anthropol.* (in the press).
10. Raff, J. A., Bolnick, D. A., Tackney, J. & O'Rourke, D. H. *Am. J. Phys. Anthropol.* **146**, 503–514 (2011).
11. Neves, W. A. & Hubbe, M. *Proc. Natl Acad. Sci. USA* **102**, 18309–18314 (2005).
12. Stanford, D. J. & Bradley, B. A. *Across Atlantic Ice: The Origin of America's Clovis Culture* (Univ. California Press, 2012).
13. Pitulko, V. V. *et al. Science* **303**, 52–56 (2004).
14. Tamm, E. *et al. PLoS ONE* **2**, e829 (2007).
15. Dillehay, T. D. *et al. Science* **320**, 784–786 (2008).
16. Misarti, N. *et al. Quat. Sci. Rev.* **48**, 1–6 (2012).
17. Achilli, A. *et al. Proc. Natl Acad. Sci. USA* **110**, 14308–14313 (2013).
18. Rasmussen, M. *et al. Nature* **463**, 757–762 (2010).
19. Raghavan, M. *et al. Nature* **505**, 87–91 (2014).
20. Olalde, I. *et al. Nature* <http://dx.doi.org/10.1038/nature12960> (2014).

COSMOLOGY

Cool start to hydrogen ionization

Simulations of the cosmos cast doubt on assumptions about the temperature of primordial hydrogen gas when it was ionized by the first stars and galaxies, complicating the interpretation of ongoing observations. SEE LETTER P.197

JUDD D. BOWMAN

Not long after the Big Bang, the Universe cooled sufficiently for electrons and protons to become bound into hydrogen atoms. Over the next few hundred million years, the primordial hydrogen gas continued to cool as the Universe expanded. For astrophysicists, the excitement starts at this point, when the first stars, galaxies and black holes are believed to have released X-rays and ultraviolet light that initially heated and later 'reionized' the primordial hydrogen. This period of reionization ended about one billion years after the Big Bang. On page 197 of this issue, Fialkov and colleagues¹ report simulations that challenge existing theoretical expectations about the final stages in this sequence.

In their simulations, the authors use improved models of X-ray emission driven by the accretion of mass in binary star systems from early cosmic epochs, and find that it may have taken much longer than previously projected to heat the primordial hydrogen between nascent galaxies. The improved models come from studies^{2,3} of the evolution of binary systems throughout cosmological time, and predict more energetic X-rays than previous models, which favoured low-energy emission. High-energy X-rays are able to travel farther before being absorbed by primordial gas. This results in a more uniform heating

of the gas, but also means that, counter-intuitively, less energy is deposited into the gas overall, owing to effects associated with the Universe's expansion.

Whereas X-rays are primarily responsible for heating the primordial gas, ultraviolet light from early stars drives its ionization. Hence, although heating is prolonged in the authors' simulations, ionization proceeds at the same pace as in previous simulations, because the model for ultraviolet emission has not changed. The result is that ionization commences before the gas is hot. This outcome is contrary to earlier expectations and has significant consequences for a new branch of observational astrophysics.

Over the past decade, astrophysicists around the world have constructed pioneering radio telescopes to detect the 21-centimetre-wavelength radio emission of neutral atomic hydrogen from the reionization era. Instruments include the Low Frequency Array (LOFAR)⁴, the Precision Array for Probing the Epoch of Reionization (PAPER)⁵ and the Murchison Widefield Array (MWA)⁶. These telescopes do not observe at 21-cm wavelengths, but rather at metre wavelengths. Owing to the Universe's expansion, 21-cm radiation from the early Universe is highly stretched (redshifted) on its way to Earth. The intensity and distribution of the redshifted 21-cm emission on the sky can be used to map the temperature, density and



50 Years Ago

The fickleness of our sensations is as old as fickleness itself. Portia tells the proud Prince of Morocco: "All that glistens is not gold, Often you have heard this told; ..."

but, no doubt, if Bacon really had written Shakespeare's plays, the scroll would have started with the lines: "Sane men say that they have seen Red turn before their eyes to green".

... While it would be folly to expect (and to claim) that all apparent colour changes are to be explained in terms of photochemical rather than nervous phenomena, it would be wise not to exclude the possibility that this may be done at least in part.
From *Nature* 15 February 1964

100 Years Ago

The issue of *The National Geographic Magazine* for January is wholly devoted to a finely illustrated article by Mr. F. E. Johnson, entitled "Here and There in Northern Africa." It contains a splendid series of photographs depicting the racial types, particularly those of the Ouled Nail dancing girls, whose performances are familiar to visitors to Biskra. The pictures of life in the harem and in the oases are very striking.

ALSO:

The properties of alcohol and of stimulants in general in relation to their physiological effects form the subject of an address, given by Prof. H. E. Armstrong to the Institute of Brewing ... The account given of the power of alcohol and its homologues, when used in moderate amounts, to penetrate the membrane which encloses the cell ... should do something to overcome the prejudice against alcohol which exists in the minds of otherwise fair-minded people.
From *Nature* 12 February 1914

ionization of the primordial gas⁷. What is more, the evolution of the gas can be traced by tuning to increasingly longer wavelengths to study gas from progressively earlier epochs (higher redshifts) in the Universe.

The first observations from these radio telescopes are emerging, but are shrouded by high levels of thermal noise and contamination from astrophysical signals other than that of the primordial gas. Observers plan to rely on simple, easily identifiable spectral features predicted by theoretical simulations to detect the neutral-hydrogen signal. One predicted feature is a rise and fall⁸ in the amplitude of fluctuations of the 21-cm signal with wavelength. The peak in amplitude is expected when roughly half of the hydrogen is ionized. Identifying the peak would pinpoint the mid-point of the reionization process — a key stage for understanding the underlying properties of the first stars and galaxies.

Fialkov and colleagues' study¹ finds more variability in this rise and fall behaviour than previously expected, together with a well-defined minimum in amplitude that should exist in the observations. The authors also predict a weaker signal overall, increasing the difficulty of observations, because more-uniform heating results in less-extreme fluctuations in the 21-cm signal. To use an analogy with mountain topography, the researchers find fluctuations that are more like the gentle curves of the foothills of the Himalayan Mountains than the jagged features of the mountains themselves (Fig. 1).

Planned telescopes currently seeking funding, such as the Square Kilometre Array (SKA)⁹ and the Hydrogen Epoch of Reionization Array (HERA)¹⁰, should have little trouble discriminating between high- and low-energy X-ray heating models if designed accordingly. Fortunately, the authors' simulation comes during a critical phase of detailed design for both instruments. The results should prompt astrophysicists to reconsider the wavelength range that the telescopes will target, and to assess the relative benefits of optimizing them for statistical measurements (as PAPER and other current instruments attempt) compared with more general direct imaging of the 21-cm signal.

Like most simulations before it, the current work relies on several approximations to make calculations tractable, particularly in modelling radiative processes. The semi-analytical methods used may miss some of the finer details, but should accurately reconstruct the big picture. Just as important as the specific outcome of the simulations, the current study highlights the inherent uncertainty in a rapidly developing field, and the often messy process of updating accepted paradigms. After a decade of tremendous theoretical advances in modelling the reionization epoch, and with few observations to provide checks, it is easy to overlook underlying assumptions. We need



Figure 1 | A topographic analogy. Previously, astrophysicists expected that relatively localized heating by low-energy X-rays would produce extreme peaks and valleys in the temperature of hydrogen gas before the gas was ionized, resembling the jagged Himalayan Mountains (background). Fialkov and colleagues¹ have used improved models of X-ray emission that yield more high-energy X-rays than previous models, and show that the gas will be cooler and fluctuations less pronounced, more like the gentle curves of the foothills in the foreground.

to remember that our current conceptions are based on models that can easily shift, and that changes to a single input may substantially alter outcomes. Fialkov and colleagues' analysis should remind us to keep our minds — and our experimental endeavours — open to a wider range of possibilities. ■

Judd D. Bowman is in the School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.
e-mail: judd.bowman@asu.edu

1. Fialkov, A., Barkana, R. & Visbal, E. *Nature* **506**, 197–199 (2014).
2. Fragos, T. et al. *Astrophys. J.* **764**, 41 (2013).
3. Fragos, T., Lehmer, B. D., Naoz, S., Zezas, A. & Basu-Zych, A. *Astrophys. J.* **776**, L31 (2013).
4. Van Haarlem, M. P. et al. *Astron. Astrophys.* **556**, A2 (2013).
5. Parsons, A. R. et al. Preprint at <http://arxiv.org/abs/1304.4991> (2013).
6. Tingay, S. J. et al. *P. Astron. Soc. Aust.* **30**, e007 (2013).
7. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. *Phys. Rep.* **433**, 181–301 (2006).
8. Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M. & Hernquist, L. *Astrophys. J.* **680**, 962–974 (2008).
9. <http://skatelescope.org>
10. <http://reionization.org>

This article was published online on 5 February 2014.

REMOTE SENSING

A green illusion

An analysis reveals that satellite-observed increases in canopy greenness during dry seasons, which were previously interpreted as positive responses of Amazon forests to more sunlight, are in fact an optical artefact. [SEE LETTER P.221](#)

KAMEL SOUDANI & CHRISTOPHE FRANÇOIS

Amazon forests cover about half of the world's area of tropical forest and constitute one of the largest reservoirs of insect, animal and plant biodiversity. The forests also have a crucial role in regulating atmospheric carbon dioxide, and thus climate-change trajectories. Nevertheless, the biogeochemical responses of these ecosystems to climate drivers remain poorly understood. For example, it is not clear whether the primary driver of the Amazon forests' seasonal patterns of greenness is the availability of light or that of water, nor how these ecosystems respond to drought. Remote-sensing studies conducted to address such questions have produced divergent results, leading to ongoing controversy. However, in this issue, Morton *et al.*¹ (page 221) present findings on the causes of satellite-observed greenness changes that could be decisive in this context.

Let us take a step back to trace the origins of the controversy. It began with studies^{2–4} reporting that Amazon forests 'green up' in the dry season. One of these studies⁴ reported an even more unexpected observation: that the forests became greener during the severe drought that occurred in the region in 2005. The authors of these studies suggested that the greenness of Amazon forests is limited by light availability and that the observed increase of

greenness during the dry season is stimulated by increased sunlight. Stunningly, they also proposed that the soil water content is not a limiting factor for greenness.

These findings were challenged by other studies^{5,6}, which concluded that the 2005 drought had no impact (positive or negative) on the greenness of Amazon forests. The authors of these papers proposed that the previous findings were attributable to artefacts resulting from contamination of satellite-based observations by clouds and aerosols. Morton and colleagues' study goes further, showing that the apparent increase in greenness in Amazon forests can be simply explained by seasonal variations in sun-sensor geometry. Their results support the hypothesis that it is seasonal soil-moisture availability — rather than light — that governs the balance between photosynthesis and respiration in Amazon forests.

To reach these conclusions, Morton *et al.* tested two categories of assumption that might explain a greenness increase during the dry season. The first category relates to temporal changes in the optical properties of leaves and in canopy structure; the second involves changes in solar illumination and sensor viewing angles at different times of year. Using a three-dimensional reconstruction of a forest and a ray-tracing model that simulates the trajectories of photons within the canopy, the authors assessed the effects of both sets of assumptions and found that

the apparent green-up of Amazon forests is driven by an increase in near-infrared reflectance during the dry season. They show that this increase is due not to changes in the optical properties of leaves or the canopy, but to a reduction in the proportion of shadow cast by canopy elements when the Sun is behind the sensor (Fig. 1).

After correcting for these angular effects, the observed increase in greenness during the dry season disappears, as do the differences in forest greenness between non-drought and drought years. Morton and colleagues confirmed these results using independent lidar satellite observations, which are not prone to sun-sensor-geometry artefacts because a laser, rather than the Sun, is used as the illuminating source. The lidar data showed no significant change in canopy attributes during the dry season. These findings clarify the current controversy and reconcile, at least partially, the conclusions of remote-sensing studies on the impacts of drought in Amazon forests with field-based observations of an increase in tree mortality during the 2005 drought⁷.

One of the great merits of Morton and colleagues' study is that it highlights the difficulties in interpreting remote-sensing data and the care that is required in doing so, especially when using such data to study phenological patterns in tropical forests. Satellite-based sensing of temporal patterns of greenness in Amazon forests is a complex challenge for several reasons. Tropical forests are composed of thousands of species, mainly broad-leaved evergreen species, which exhibit small and subtle temporal phenological variations. The Amazon region is also very cloudy during the rainy season and is frequently covered by shallow cumulus clouds⁸ even during the dry season, a phenomenon that is strongly correlated with local evapotranspiration⁹. These atmospheric conditions considerably reduce the number of cloud-free images that can be acquired, and even pixels classified as 'cloud free' in satellite-based data may be corrupted by clouds, fog or smoke from biomass burning¹⁰. There are also known inconsistencies between optical-greenness indices acquired from different sensors, products and collections¹¹.

More importantly, there is a cruel lack of a long-term ground-based observational network for the validation and calibration of remote-sensing data. All of these factors make it difficult to use satellite-based methods to accurately detect temporal changes in Amazon forest greenness. Morton and colleagues' study has rung alarm bells about the use of remote-sensing observations in this context, and highlights the need for 'ground truths' from Amazon forests. ■

Kamel Soudani and Christophe François
are in the Ecology, Systematic and Evolution
Laboratory (CNRS–University of Paris-Sud),

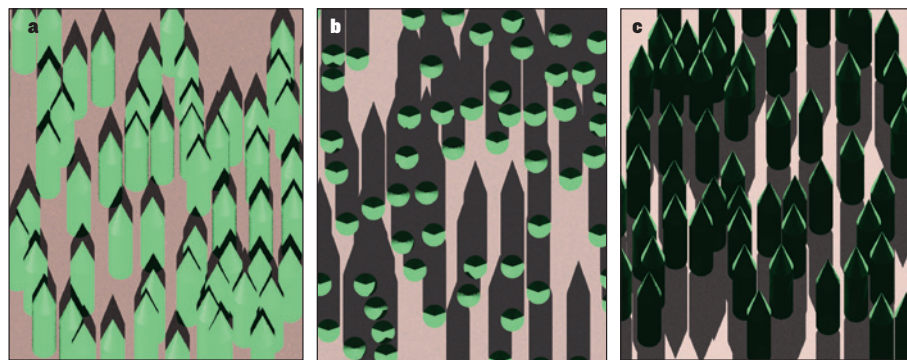


Figure 1 | Viewing angle. Morton *et al.*¹ show that the apparent increase of greenness in Amazon forests during the dry season is an optical illusion created by the relative position of the Sun during this season. The effect is exemplified by these images produced using ray-tracing software and a bi-directional reflectance distribution function model. **a**, When the Sun is in the south, behind an observer (or a remote-sensing device) also in the south, the observer sees minimal tree shadow, so the forest looks brighter and more 'green'. **b**, The view if the Sun is still in the south but the observer is right above the forest (the nadir direction). **c**, If the observer is in the south but the Sun is in the north, the shadow seen by the observer is maximized and the forest looks darker. (Images provided by Sylvain Leblanc, Natural Resources Canada.)

91400 Orsay, France.

e-mails: kamel.soudani@u-psud.fr;
christophe.francois@u-psud.fr

1. Morton, D. C. *et al.* *Nature* **506**, 221–224 (2014).
2. Huete, A. R. *et al.* *Geophys. Res. Lett.* **33**, L06405 (2006).
3. Myneni, R. B. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 4820–4823 (2007).
4. Saleska, S. R., Didan, K., Huete, A. R. & da Rocha, H. R. *Science* **318**, 612 (2007).
5. Samanta, A. *et al.* *Geophys. Res. Lett.* **37**, L05401 (2010).
6. Samanta, A., Ganguly, S., Vermote, E., Nemani, R. R.

& Myneni, R. B. *Environ. Res. Lett.* **7**, 024018 (2012).

7. Phillips, O. L. *et al.* *Science* **323**, 1344–1347 (2009).
8. Koren, I., Kaufman, Y. J., Remer, L. & Martins, J. V. *Science* **303**, 1342–1345 (2004).
9. Heiblum, R. H., Koren, I. & Feingold, G. *Atmos. Chem. Phys. Disc.* **13**, 30013–30037 (2013).
10. Samanta, A., Ganguly, S., Vermote, E., Nemani, R. R. & Myneni, R. B. *Earth Interact.* **16**, 1–14 (2012).
11. Huete, A. R. & Saleska, S. R. in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. XXXVIII, Part 8, 539–541 (Copernicus, 2010).

This article was published online on 5 February 2014.

CONDENSED-MATTER PHYSICS

History matters for a stirred superfluid

The observation of path dependence in the response of a superfluid to stirring promises potential applications in precision rotation sensing, and provides a test bed for microscopic theories of ultracold atomic gases. [SEE LETTER P.200](#)

MATTHEW J. DAVIS & KRISTIAN HELMERSON

Hysteresis is the dependence of a system on not only its present surroundings, but also its history. It has been observed in many branches of science and engineering¹, such as in the elasticity of materials, electronic circuits, biological processes, and even in macroeconomics. In many of these systems, hysteresis can be exploited for applications. On page 200 of this issue, Eckel *et al.*² at the Joint Quantum Institute in Gaithersburg, Maryland, report the observation of hysteresis in the superfluid flow of an ultracold cloud of atoms known as a Bose–Einstein condensate. This result could prove an important step for developing atomtronic devices, in which atoms have an equivalent role to that of electrons in electronic devices. It could also provide a platform

for investigating the microscopic origins of hysteresis and of the dissipative dynamics of Bose–Einstein condensates.

Perhaps the best-known example of hysteresis is the response of the magnetization of a ferromagnetic material to an applied magnetic field. If the material is initially unmagnetized, slowly turning on a magnetic field will cause the material to become magnetized. However, subsequently reducing the magnetic field back to zero will leave the system with a net magnetization — the system is now in a different state from that in which it started, despite the environment being identical.

More generally, systems that exhibit hysteresis consist of many interacting particles, allowing a complex myriad of states. Although the constituent particles and their interactions can be described by equations that are fully

reversible in time, their collective behaviour is not. For example, for a single atom, the direction of its magnetic moment depends only on the applied magnetic field at that instant in time. However, in a ferromagnet, the collective behaviour of the atoms' magnetic moments, which involves the formation of magnetic domains of uniform magnetization, does not. This hysteretic effect is the basis of modern magnetic-recording media such as hard disks.

Superfluidity and superconductivity are macroscopic quantum phenomena that generally arise owing to the presence of a Bose–Einstein condensate (BEC), in which a huge number of identical interacting particles shares the same quantum-mechanical wavefunction. These quantum liquids are 'super' because, below a critical velocity, they can flow past obstacles without any friction or resistance³.

When a BEC is confined to a toroidal (doughnut-shaped) container, the consequences are particularly interesting. One of the mathematical requirements of the wavefunction, which is characterized by an amplitude and a phase, is that it must be single-valued and continuous, and this means that the total change of a wavefunction's phase about the doughnut must be an integer multiple of 2π . Quantum mechanics dictates that the speed of the particles in the ring is proportional to the rate of change of the phase about the container, which means that the circulation (the integral of the speed around the ring) of the superfluid is restricted to integer multiples of a constant — the circulation is quantized. Flows with speeds below the critical velocity last indefinitely, and are known as persistent currents³. For many researchers, this is the main feature of superfluidity and superconductivity.

The team at the Joint Quantum Institute has been working with BECs in ring-shaped traps for the past few years, and has created and observed the persistence of flows⁴. The group has studied⁵ the decay of persistent flows following the introduction to the ring of an obstacle formed by a laser beam, creating a constriction (weak link) that the fluid

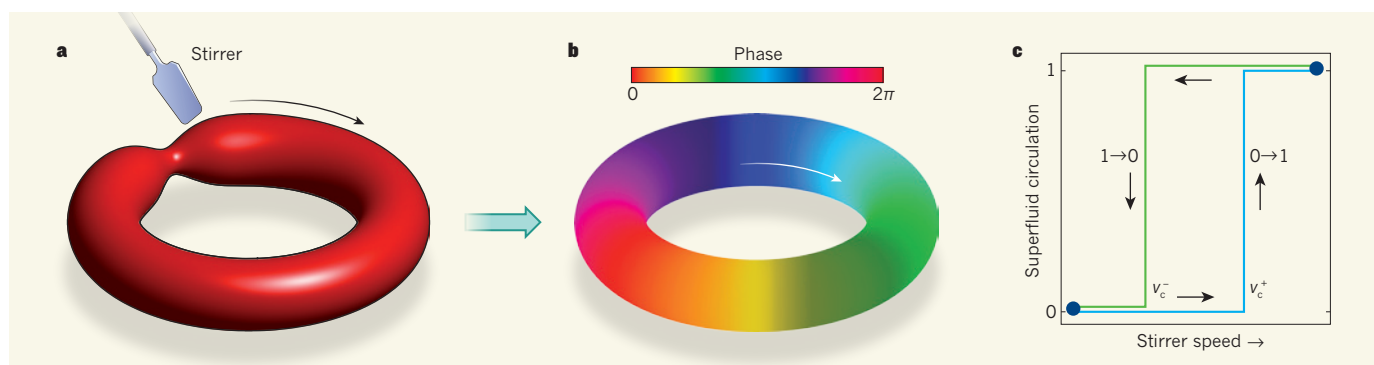


Figure 1 | Hysteresis in a stirred superfluid². **a**, A dilute-gas superfluid with a constant quantum-mechanical phase is initially held at rest in a ring-shaped trap. An obstacle rotating about the ring is then introduced, creating a constriction and stirring the superfluid. **b**, Above a critical stirrer speed (v_c^+), the stirring results in a transition from zero (0) to one (1) unit of circulation (the integral of the speed around the ring) in the superfluid, and the final phase of the system varies continuously from zero to 2π about the loop. Below v_c^+ , the superfluid does not respond to the stirring. **c**, The transition from 0 to 1 occurs as the stirrer speed increases past v_c^+ . However, on reversing the path, the transition from 1 to 0 takes place at the lower speed v_c^- . This cyclic path is an idealized example of hysteresis.

must pass through. More recently, the team has used⁶ the obstacle as a paddle-like stirrer to induce transitions from a state with no circulation to one with one or more quanta of circulation.

In the current study, Eckel and colleagues begin with a superfluid at rest in the ring trap. The researchers turn on a stirrer rotating at constant speed, and leave it in the superfluid for two seconds (Fig. 1a). The stirrer is then withdrawn and the effect it has had on the superfluid is observed. For stirring speeds below a critical value, the authors find that the superfluid is oblivious to the stirring — it remains at rest in a state with no circulation. However, above this critical value, they find that the fluid transitions to a state in which it moves about the ring with one quantum of circulation (Fig. 1b).

Next, Eckel *et al.* run their experiment in reverse: they begin with the superfluid rotating with one quantum of circulation. They stir it over the same range of speeds and directions as before, but because the stirrer is now moving more slowly than the superfluid, they look for transitions to the state in which the superfluid is at rest. Their key observation is that the stirring speed that returns the superfluid to rest can be smaller than the speed required to cause it to circulate in the first place — an example of hysteresis (Fig. 1c).

Why is this result interesting? The researchers' system is an atomtronic device analogous to a radio-frequency superconducting quantum interference device (SQUID) — a superconducting loop with one weak link. SQUIDs can exhibit hysteretic behaviour, and have applications in sensing vanishingly small magnetic fields. Owing to the mathematical analogy between magnetic fields for charged particles and rotation for neutral particles, atomtronic SQUIDs could prove to be exceptional rotation sensors. Although other researchers have demonstrated atomtronic SQUIDs^{7,8}, Eckel *et al.* are the first to observe hysteresis in such a system. The authors suggest that this effect could play a crucial part in the design and application of atomtronic circuits, as it has in other electronic devices. Whether this is true remains to be seen. The level of control obtained during the experiment is impressive, and the data are remarkably clean. However, for atomtronic devices to prove genuinely useful and move out of the research lab, their reliability and robustness needs to continue to improve.

Eckel and colleagues also simulate the stirring of the superfluid using the Gross–Pitaevskii model of BECs, which has proved remarkably adept at predicting the outcome of many experiments on dilute-gas superfluids³. And this is where the authors' study gets particularly interesting. Although the simulations qualitatively reproduce the hysteretic behaviour, they predict critical stirring speeds that are too large by at least a factor of two.

This provides a real challenge for theorists — what is going wrong with the model? One issue is that the model does not allow for dissipation of energy, which is essential for hysteresis¹. However, even adding dissipation to the model hardly changes the predictions. This result could perhaps have been anticipated from theoretical studies⁹ of an earlier experiment⁵ performed at the Joint Quantum Institute that found evidence of physics beyond the Gross–Pitaevskii model⁹ in the decay of persistent currents following the introduction of an obstacle. The team proposes a simple model of the energetics of vortex-pair creation in the presence of the stirrer that agrees rather well with their experimental results. It is up to theorists to try to work out whether this model is correct, and how it emerges from the full microscopic theory of BECs.

Finally, hysteresis in real materials is often described by phenomenological models¹. The complexity of these multi-particle systems generally precludes a description of hysteresis from first principles. In atomic BECs, in which the microscopic interactions between the atoms are well understood³ and the number of constituent particles is computationally tractable, it may

be possible to go beyond the Gross–Pitaevskii model and discern the ingredients needed to produce hysteresis and further exploit it for technological applications — the past can guide us towards a brighter future. ■

Matthew J. Davis is at the School of Mathematics and Physics, University of Queensland, St Lucia, Queensland 4072, Australia. **Kristian Helmerson** is at the School of Physics, Monash University, Clayton, Victoria 3800, Australia.
e-mail: mdavis@physics.uq.edu.au

1. Mayergoyz, I. D. *Mathematical Models of Hysteresis and their Applications* 2nd edn (Academic, 2003).
2. Eckel, S. *et al.* *Nature* **506**, 200–203 (2014).
3. Leggett, A. J. *Quantum Liquids: Bose Condensation and Cooper Pairing in Condensed-Matter Systems* (Oxford Univ. Press, 2006).
4. Ryu, C. *et al.* *Phys. Rev. Lett.* **99**, 260401 (2007).
5. Ramanathan, A. *et al.* *Phys. Rev. Lett.* **106**, 130401 (2011).
6. Wright, K. C., Blakestad, R. B., Lobb, C. J., Phillips, W. D. & Campbell, G. K. *Phys. Rev. Lett.* **110**, 025302 (2013).
7. Ryu, C., Blackburn, P. W., Blinova, A. A. & Boshier, M. G. *Phys. Rev. Lett.* **111**, 205301 (2013).
8. Sackett, C. A. *Nature* **505**, 166–167 (2014).
9. Mathey, A. C., Clark, C. W. & Mathey, L. Preprint at <http://arxiv.org/abs/1207.0501> (2012).

CONSERVATION

Making marine protected areas work

Globally consistent surveys of five factors influencing the success of marine protected areas — age, size, isolation, protection and enforcement — reveal that only when all five are present does nature thrive. SEE LETTER P.216

BENJAMIN S. HALPERN

In the past few years, several huge marine protected areas (MPAs) have been created in the Pacific and Indian oceans, totalling more than 1.6 million square kilometres. That might sound like a large area, but even when combined with every other MPA on the planet, still less than around 2% of the world's oceans are fully protected. In response to this conservation shortfall, coastal nations have committed to increasing the amount of protected area in their territorial waters to at least 10% by 2020. However, even if this target is met, the actual conservation value may be limited because MPAs often exist in name only — they do not truly provide protection. On page 216 of this issue, Edgar *et al.*¹ provide key insights into why so many of the world's MPAs are failing to meet their full potential and describe a clear path forward for achieving better conservation outcomes.

The problems with existing protected areas are manifold. 'Paper parks', for which

protected-area boundaries exist in principle but are not enforced, have little conservation value. Unfortunately, there are a lot of paper parks^{2,3}, and many other protected areas are only partially protected — activities such as recreational or hook-and-line fishing are allowed, leading to less conservation value than full protection⁴. Furthermore, nearly half of all MPAs are little bigger than a football field or have only recently been created⁵, limiting their ability to protect many species.

It might seem as if we know a lot about what leads to MPA success or failure, but the simultaneous assessment of how various factors affect MPA success has been missing from previous studies. In other words, most studies tested one factor at a time, without controlling for the others. This is what makes Edgar and colleagues' work particularly unusual. Using data from 87 MPAs around the world, all sampled with the same methods, the authors compared how the biomass, abundance and diversity of species in MPAs varied with all

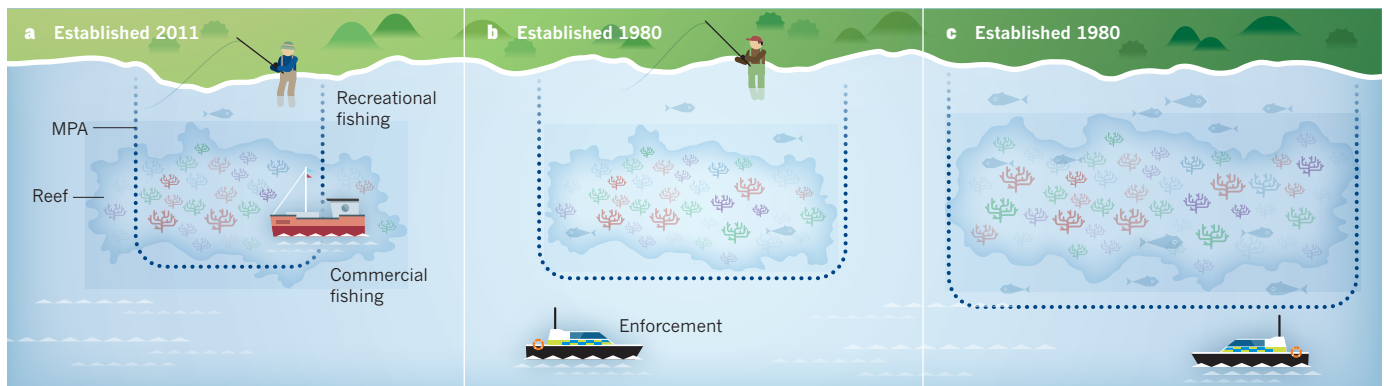


Figure 1 | A topographic analogy. Previously, astrophysicists expected that relatively localized heating by low-energy X-rays would produce extreme peaks and valleys in the temperature of hydrogen gas before the gas was ionized, resembling the jagged Himalayan Mountains (background). Fialkov and colleagues¹ have used improved models of X-ray emission that yield more high-energy X-rays than previous models, and show that the gas will be cooler and fluctuations less pronounced, more like the gentle curves of the foothills in the foreground.

combinations of five factors: whether or not the zones were fully protected, enforced, old (more than 10 years), large (more than 100 km²) and isolated (for example, a remote island or a patch of reef surrounded by large sandy areas).

Their findings are remarkable and unsettling. Instead of each factor being important, such that benefits accrued proportionally to the number of factors present, the authors found that conservation value was nearly absent in MPAs with fewer than three of these factors, and then increased exponentially from three to five factors (Fig. 1). This result is worrying because it is difficult for managers to achieve all five factors.

The authors also report a secondary, but equally interesting, result. As a by-product of the methods used to determine MPA effectiveness, Edgar *et al.* were able to estimate that fish biomass in coastal regions around the world has, on average, been reduced by about two-thirds relative to pristine conditions. Given the difficulty in assessing near-shore, small-scale fisheries in most parts of the world, this result and its underlying methods should help to improve those stock assessments and thus fisheries management, especially in developing nations.

A key concern about Edgar and colleagues' study is that only 4.6% of the assessed MPAs had all five features, and only an additional 5.7% had four. This is a limitation of the real world — only a few MPAs exist that achieve more than three of the factors — but, nonetheless, it creates a potential bias in their results. Clearly, more research is needed to better understand the generality of the authors' results, and how other factors could influence conservation success. In particular, the past decade has seen advances in the assessment and design of networks of MPAs⁶, and such research suggests that the connectivity provided by networks might compensate for other missing factors, especially small size and lack of isolation. Most managers do not have the option of placing protected areas

in isolated locations (the most influential factor, according to this study) or creating large reserves, such that the menu of design options is constrained to a set of variables that can at best achieve suboptimal results. Future work needs to identify how network connectivity and other MPA attributes might help to achieve maximum conservation success.

Limitations aside, this study is an important advance for several reasons. It is global in extent and consistent in methodology, mitigating the common critique that large-scale studies are not relevant to specific sites or comparable from location to location. The value of this aspect of the study cannot be overstated. The findings also help to calibrate expectations for existing and future MPAs. It is clear that designating and enforcing park boundaries, although necessary, is not sufficient to gain full conservation benefits, and that protected areas without all five features should not be expected to produce

such benefits. The authors frame these lessons in the context of international conservation objectives, but they are arguably more important for local managers and stakeholders. ■

Benjamin S. Halpern is in the Bren School of Environmental Science & Management, University of California, Santa Barbara, Santa Barbara, California 93106, USA, and in the Faculty of Natural Sciences, Imperial College London, UK.
e-mail: halpern@bren.ucsb.edu

1. Edgar, G. J. *et al.* *Nature* **506**, 216–220 (2014).
2. Mora, C. *et al.* *Science* **312**, 750–751 (2006).
3. Jameson, S. C. *et al.* *Mar. Pollut. Bull.* **44**, 1177–1183 (2002).
4. Lester, S. E. & Halpern, B. S. *Mar. Ecol. Prog. Ser.* **367**, 49–56 (2008).
5. Wood, L. J. *et al.* *Oryx* **42**, 340–351 (2008).
6. Gaines, S. D. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 18251–18255 (2010).

This article was published online on 5 February 2014.

OCEAN SCIENCE

Eddy effects on biogeochemistry

Ocean eddies tens of kilometres in radius can delineate local ecosystems and contribute to biogeochemical budgets. The characterization of three such eddies in a coastal upwelling region provides insight into these wonders of nature.

AMALA MAHADEVAN

The eastern margins of oceans are exceptional regions. The alongshore winds drive a mean offshore flow of surface water that creates upwelling at the coast (Fig. 1). This draws nutrient-rich, cold water from the deeper layers into the sunlit (euphotic) zone, where it elicits the rapid growth of phytoplankton, resulting in highly

biologically productive seas. The alongshore currents are unstable, and shed eddies and filaments of cold water 10–100 kilometres in length, which transport water offshore. These dynamic features affect the primary productivity of the regions, the fate of particulate organic carbon produced by photosynthesis, and the regions' biogeochemical budgets. Writing in *Biogeosciences*, Stramma *et al.*¹ provide a detailed biogeochemical perspective of

three such eddies off the coast of Peru.

The eddies surveyed extended several hundred metres in depth and exceeded 100 km in diameter. The authors found that these eddies trapped waters at the time of eddy formation, and propagated westward as biogeochemically anomalous water masses for several months. As the eddies matured, the ecosystems sustained within them modified the properties of the trapped water, consuming nutrients (nitrate and phosphate). However, the ratio of nitrogen to phosphorus in the eddies was lower than is usually seen elsewhere. This suggests that the eddies harbour nitrate-reducing (denitrifying) bacteria typical of low-oxygen environments.

Stramma and colleagues observed a range of characteristics for the individual eddies. The two anticyclonic eddies (those having a rotation opposite to that of Earth) were low in nitrate, a nutrient used by phytoplankton, but high in nitrite — which was possibly produced as an intermediate of denitrification or secreted by phytoplankton². The concentration of chlorophyll beneath the surface ‘mixed’ layer in these eddies was higher than in surrounding waters, which might indicate that they contribute positively to productivity. There were differences between the two eddies, however: phosphate was much deeper in the older of the pair. One explanation for this is that, as the older eddy propagated westward, some of its nutrients subducted without being consumed by phytoplankton, thereby suppressing productivity in the eddy compared with that of surrounding waters, in which phosphate seems to be more fully used. By contrast, the cyclonic eddy had elevated nitrate levels in its core, but no nitrite or signs of denitrification.

The authors’ study reflects a continuing interest in the effects of fronts (boundaries between waters of different density at which currents are intensified) and eddies, which are often formed by the detachment of frontal meanders. Eddies can persist for weeks, harbouring productive ecosystems³ and generating anomalous biogeochemistry⁴. Fronts are highly dynamic and generate filaments — water streamers that infiltrate from one side of a front to the other and have a different density from their surroundings. Filaments transport cold, dense water offshore in coastal upwelling zones (Fig. 1). Fronts and eddies enhance primary productivity in areas lacking surface nutrients by inducing an upward supply of nutrients into the euphotic layer⁵. But eddies and filaments may reduce biological production⁶ in coastal upwelling regions — such as those off Peru — because they contribute to the export (the lateral and downward transport) of phytoplankton-rich waters, whose nutrients are not fully consumed.

The potential suppression of phytoplankton production by eddies and filaments in upwelling ocean margins can be explained as follows. Assuming the existence of a steady

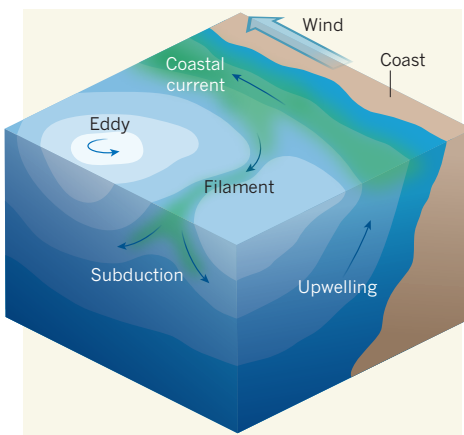


Figure 1 | An eastern upwelling ocean margin in the Southern Hemisphere. Alongshore winds heading towards the Equator transport surface water offshore (not shown) and generate upwelling at the coast. A coastal current, also heading towards the Equator, is created by the contrast in density between cold upwelled water at the coast (darker shades of blue) and warmer surface waters offshore (lighter shades of blue and grey). The upwelled water is nutrient rich and causes phytoplankton (green) to grow close to the surface. The coastal current is unstable and generates eddies, and the denser water that is upwelled nearshore is drawn offshore in filaments. The filament of colder, denser water is connected to (colder) subsurface water of similar density. As it is drawn offshore, downwelling occurs on the inner flanks of the filament, exporting phytoplankton carbon, nutrients and oxygen to the subsurface layers along isopycnals (surfaces of constant density). (Figure adapted from a graphic by M. M. Omand.)

state when averaging over time and space over the sunlit layers of an upwelling region, the levels of nutrients consumed by phytoplankton production (P) and nutrients exported (E) are balanced by the levels of nutrients supplied (S) to the upwelling region: $S = P + E$. Thus, if eddies and filaments that have formed from an upwelling front increase E , but do not affect S , which is driven by large-scale coastal upwelling, then P must decrease. If, on the other hand, eddies and fronts increase S , as occurs in subtropical gyres, then they lead to an increase in P .

The high primary productivity in upwelling regions generates particulate organic carbon of various forms. Phytoplankton pass through the food web, resulting in detritus that is heavier than sea water, and sinks. As it sinks, this organic matter decomposes, using up oxygen. At the surface, dissolved oxygen is replenished by photosynthesis and tends to equilibrate with the atmosphere. But at greater depths, eastern upwelling ocean margins are anomalously low in oxygen because the rate of oxygen resupply is insufficient to meet the high demand for organic-matter decomposition.

Eddies and filaments can influence the oxygen supply to these oxygen-poor layers beneath the surface. Filaments cause the subduction of productive phytoplankton-

rich surface waters (Fig. 1), but this export of carbon-containing material differs in a crucial way from the sinking of particulate organic carbon. The downwelling water carries with it high levels of oxygen from the surface layer, and thus provides the necessary oxygen for decomposition of the organic carbon. In this way, eddies and filaments support carbon export without placing any demand on the oxygen of subsurface waters. Without the downward transport of oxygen by eddies and filaments, the subsurface waters in upwelling zones would be even more starved of oxygen.

Eddies also have a role in the transport of momentum, heat and mass. The physical structure of eddies off Peru was recently studied using three-dimensional modelling⁷, ship-based measurements⁸ and satellite observations^{8,9}, revealing an equal propensity for cyclonic and anticyclonic eddies to develop in this region. The transport of colder waters in the eddy cores made no contribution to the heat budget in the region⁸, but whether the transport of nitrogen and carbon by these eddies affects biogeochemical budgets is difficult to assess from the measurements available so far.

The diverse characteristics of the three eddies sampled by Stramma *et al.* show clearly that assessing the integrated contribution of eddies to the primary productivity, carbon export and biogeochemistry of upwelling regions will require a much more statistically complete characterization of properties within eddies, filaments and their surroundings. Sustained time series and autonomous biogeochemical sampling (for example, see ref. 10) are needed, in concert with physical measurements at high resolution in space and time, to capture the episodic transport and biogeochemical transformations that occur in the dynamic and biogeochemically important coastal upwelling margins of the world. ■

Amala Mahadevan is in the Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.
e-mail: amala@whoi.edu

1. Stramma, L., Bange, H. W., Czeschel, R., Lorenzo, A. & Frank, M. *Biogeosciences* **10**, 7293–7306 (2013).
2. Collos, Y. *Mar. Ecol. Prog. Ser.* **171**, 293–301 (1998).
3. McGillicuddy, D. J. Jr *et al.* *Science* **316**, 1021–1026 (2007).
4. Benitez-Nelson, C. R. & McGillicuddy, D. J. Jr *Deep-Sea Res. II* **55**, 1133–1138 (2008).
5. Mahadevan, A. & Archer, D. J. *Geophys. Res.* **105**, 1209–1225 (2000).
6. Gruber, N. *et al.* *Nature Geosci.* **4**, 787–792 (2011).
7. Colas, F., McWilliams, J. C., Capet, X. & Kurian, J. *Clim. Dynam.* **39**, 509–529 (2012).
8. Holte, J., Straneo, F., Moffat, C., Weller, R. & Farrar, J. T. *J. Geophys. Res.* **118**, 2295–2309 (2013).
9. Chaigneau, A., Le Texier, M., Eldin, G., Grados, C. & Pizarro, O. *J. Geophys. Res.* **116**, C11025 (2011).
10. Johnson, K. S. *et al.* *J. Atmos. Oceanic Technol.* **30**, 1854–1866 (2013).

This article was published online on 29 January 2014.

Cold dark matter heats up

Andrew Pontzen^{1,2,3} & Fabio Governato⁴

A principal discovery in modern cosmology is that standard model particles comprise only 5 per cent of the mass–energy budget of the Universe. In the Λ CDM paradigm, the remaining 95 per cent consists of dark energy (Λ) and cold dark matter. Λ CDM is being challenged by its apparent inability to explain the low-density ‘cores’ of dark matter measured at the centre of galaxies, where centrally concentrated high-density ‘cusps’ were predicted. But before drawing conclusions, it is necessary to include the effect of gas and stars, historically seen as passive components of galaxies. We now understand that these can inject heat energy into the cold dark matter through a coupling based on rapid gravitational potential fluctuations, explaining the observed low central densities.

Despite the unknown nature of the dominant components, the Λ CDM paradigm successfully describes the evolution of the Universe from its near-uniform early state, as measured by the cosmic microwave background¹, to the present-day clustered distribution of matter^{2,3} in an accelerating Universe. Consequently, the properties of dark matter and the processes driving the formation and evolution of galaxies are fundamental, closely connected problems in modern astrophysics.

Λ CDM, through its explanation of observations on the largest observable scales, has been established as the standard cosmological paradigm. Over time, increasingly massive dark matter ‘halos’ form through gravitational instabilities, starting from small perturbations in the matter density field. It is within the gravitational potential of dark matter halos that galaxy formation—gas cooling and star formation—proceeds⁴. However, long-standing problems have been encountered in reconciling the predictions of Λ CDM with observational results at galaxy scales. These problems probably stem from our poor understanding of the complex physics associated with star formation, and are complicated by our failure to identify the dark matter particle.

Recently, major progress has been made in addressing these discrepancies. We now understand that gas outflows from galaxies are ubiquitous, powered by energy released from stars and black hole accretion. These outflows change the distribution of the gas and stars that subsequently form. If the outflows launch at sufficient speed, they also cause an irreversible change in the dark matter distribution, even if the gas later returns to the galaxy in a ‘fountain’. The mechanism is purely gravitational and therefore independent of the nature of the particle. These processes fundamentally modify the structure of galaxies, and serve to bring theoretical expectations into agreement with previously problematic observational constraints. These astrophysical processes must therefore be fully understood before attempting to use galaxies to falsify Λ CDM.

Galaxy formation with cold dark matter

The viability of the Λ CDM picture of structure formation was first evaluated using computer simulations (allowing, for instance, neutrinos to be ruled out as the dominant component of dark matter⁵). Gas cooling and star formation within dark matter halos is now the standard paradigm for the origin of galaxies⁴. The behaviour of dark matter can be simulated on computers by chunking a portion of the universe into ‘particles’ and programming suitable physical laws to dictate how these particles move over time. Because simulated particles typically interact only through gravity, such simulations are called collisionless.

Early attempts used just 30,000 particles to follow large regions of the Universe. Consequently each particle had the mass of a large galaxy; even so, such simulations were expensive, taking 70 hours on state-of-the-art 3-MHz facilities. Such calculations would now take a few minutes on a mobile telephone. The growth of computing power and parallel capabilities meant that, by the 1990s, simulations became sufficiently powerful to make detailed predictions of the internal structure of halos in different cosmological scenarios. These simulations highlighted the universal nature of dark matter halos formed through collisionless collapse. The spherically averaged density of halos is ‘cusped’ at the centre (scaling approximately as $\rho \propto r^{-1}$, where ρ is the halo density and r is the radius from its centre), rolling to a steeper slope at larger radius (reaching $\rho \propto r^{-3}$); such behaviour is known as ‘NFW’, after the authors of a pivotal paper⁶.

At the same time, simulations started highlighting a number of deficiencies in the cold dark matter scenarios. The most evident was the overabundance, by more than an order of magnitude, of small satellites^{7,8} compared to the number observed orbiting the Milky Way⁹ at the time. Worse, the simulations significantly over-predicted the density of dark matter at the centre of galaxies^{10,11}. Increasingly precise observations of the rotation curves of field galaxies have confirmed this discrepancy¹² (see below).

Collisionless dark matter simulations have since reached maturity, with modern simulations using several billion resolution elements for just one Milky-Way-sized halo^{13,14}. However, to make predictions that are testable against observations of the real Universe, baryon physics must be introduced. (Here we are adopting the astronomical convention of referring to baryons and leptons collectively as ‘baryons’.) Because baryons dissipate energy and so collapse to smaller scales than dark matter, they constitute a sizeable fraction of the mass in the central regions of all but the faintest galaxies¹⁵. Moreover, observational constraints on galaxy formation ultimately come from photons, which can only be produced by baryons. Accordingly, much effort has recently been devoted to implementing gas hydrodynamics and a description of star formation within simulations^{16–19}.

The energy released by young stellar populations and active galactic nuclei into the surrounding intergalactic medium is critical for regulating star formation⁴. Without this energy, most of the gas becomes cold and dense, rapidly collapsing to form stars, contradicting observations. Processes providing the energy to halt collapse are collectively named ‘feedback’ and include supernova winds, radiation from young stars, and radiation and heat from black hole accretion^{20–23}. Including these effects

¹Department of Physics and Astronomy, University College London, London WC1E 6BT, UK. ²Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. ³Balliol College, University of Oxford, Broad Street, Oxford OX1 3BJ, UK. ⁴Astronomy Department, University of Washington, Seattle, Washington 98195, USA.

has led to strides forward in forming realistic disk galaxies, reproducing the efficiency of star formation as a function of total galaxy mass, and linking gas accretion and mergers to galaxy morphology^{24–26}. However, until recently any direct effect of the baryonic component on the dark matter was limited to a minor ‘adiabatic’ correction²⁷. In this picture, the added gravitational attraction of gradually accumulating gas causes the dark matter to contract slightly. Star-formation processes resulted in ‘passive’ changes to the galaxy population—modulating the star-formation rate without significant changes to the underlying cosmic dark matter scaffolding.

This picture has recently been subverted. Spectroscopic observations reveal the ubiquity of massive galaxy outflows driven by feedback, carrying significant gas mass away from star-forming galaxies throughout cosmic history^{28–30}. It has slowly been realized that these directly observed processes have a non-adiabatic impact on the associated dark matter halos. The effect is to relieve discrepancies between baseline cold dark matter simulations and the real Universe. Our emerging understanding of these processes constitute the central part of this Review.

Evidence that galactic gas is expelled

There is clear observational evidence that star-formation activity drives gas out of galaxies (Fig. 1). This largely arises from studies of the resonance absorption lines imprinted into spectra by the presence of heavy elements. Consequently, dramatic advances in our knowledge have been made possible by 10-m-class telescope spectroscopy with instruments including the Keck telescope’s DEIMOS²⁹ and LRIS³¹. One can either look for blueshifted absorption in the spectra of galaxies themselves^{30,32} or at ‘intervening’ features imprinted on the spectra of more distant background light sources³¹. A natural source for the energy required to generate these outflows are supernovae³³ and ionizing radiation³⁴ associated with stellar populations. In addition, energy released during accretion onto a massive central black hole may have a role, although the

available energy is thought to scale steeply with the black hole’s mass, limiting these effects to the brightest galaxies or their progenitors^{20,21}.

Recent results underline the ubiquity of outflows³⁵ from even the smallest objects (Fig. 1). Galaxies are surrounded by metal-enriched gas moving at hundreds of kilometres per second³⁶ in bubbles extending to a hundred kiloparsecs or more. This result is exceptionally hard to explain without significant galactic winds. Mounting evidence also suggests that much of the in-flowing material into galaxies may also be metal-enriched³⁷, consistent with a picture in which much of the wind does not attain the escape velocity but instead re-accretes³⁸.

A separate argument also points to the importance of winds during galaxy formation. Observed stellar profiles of small galaxies are mostly ‘bulgeless’, that is, well approximated by a disk of gas and stars with an almost exponential profile^{39,40}. Yet cosmological simulations show that the dark matter and baryons accumulated in all galaxy halos contain a large fraction of low-angular-momentum material⁴¹—which would imply the presence of a bulge⁴². This problem, known as the ‘angular momentum catastrophe’, is solved if low-angular-momentum gas is ejected^{43,44} by winds at relatively high redshift, when star formation peaks⁴⁵. Hence, the physics of galactic winds is of fundamental importance to understanding the population of disk galaxies, even before the effect on dark matter is considered.

Evidence for a cusp–core discrepancy

We now turn our attention to the excessive quantity of dark matter predicted by the cold dark matter model compared to measured densities in the innermost regions of galaxies.

Dwarf galaxies

As explained above, the under-abundance of dark matter in the centre of dwarf galaxies relative to theoretical predictions is known as the cusp–core discrepancy. The problem was discovered as soon as cosmological simulations became capable of predicting halo structure^{46,47}. Although acceptance was gradual, it is now firmly established that robust measurements of the dark matter density can be made from the rotation curves of gas-rich dwarf galaxies ‘in the field’ (that is, away from the influence of larger galaxies). In the innermost regions (of radius less than about half a kiloparsec) the baryonic contribution to the potential is comparable to that of the dark matter and must be subtracted^{12,48}. Consequently, inferring the dark matter density requires (1) high spatial resolution of the gas and stellar kinematics, (2) a comprehensive understanding of how to estimate and subtract the stellar and gas mass distribution from the central kiloparsec and (3) careful handling of systematic observational errors. The last category encompasses possible biases arising from radio beam-smearing, departure from circular orbits, centring difficulties, unknown details of stellar mass-to-light ratios and gravitational potential asphericity within galaxies; these are now thought to be under control, given that we can test algorithms on mock observations from simulations (where the true density is known)^{49–51}.

Results from recent surveys of the local Universe, such as THINGS and LITTLE THINGS^{52,53}, can therefore be regarded as free from significant observational bias. These samples reveal shallower-than-NFW dark matter profiles in a large fraction of dwarf field galaxies, with $\rho \propto r^{-0.4}$ interior to the central kiloparsec (Fig. 2). The objects are referred to as ‘cored’ although the estimated density profile is almost never actually flat. So, after 20 years of study, the cusp–core problem has remained a persistent and significant discrepancy between theoretical models of a Λ CDM universe and observations of dwarf galaxies.

Milky Way satellites

Small galaxies known as ‘dwarf spheroidals’ orbit close to the Milky Way. These dwarf spheroidals have little gas content and their stellar content is not in a rotational disk³. This probably reflects the effect of tidal fields and strong interactions with the hot gas in the halo of the parent galaxy⁵⁴. Sampling the smallest halo masses in which galaxies form, these satellites have the potential to constrain the properties of

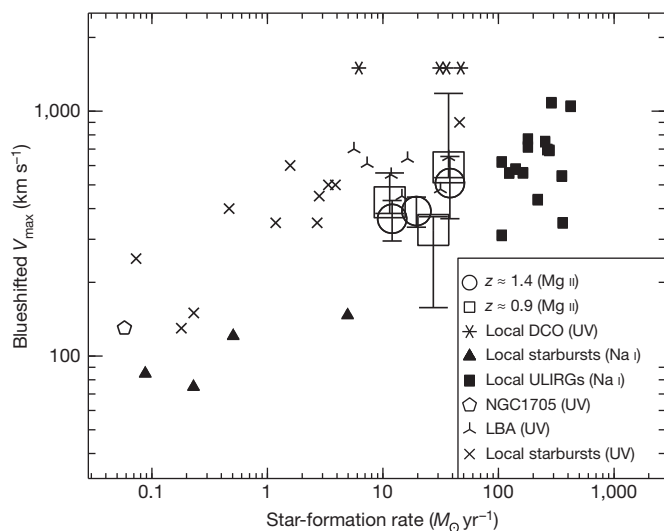


Figure 1 | Outflowing gas is ubiquitous around galaxies. Here, the absorption line blueshifts for cool gas are shown as a function of the galaxy’s star-formation rate³⁰. Even dwarf galaxies with star-formation rates under $1 M_{\odot} \text{ yr}^{-1}$ are able to support winds exceeding 100 km s^{-1} . The outflow rate of these winds is typically several times the instantaneous star-formation rate of the parent galaxy. Smaller symbols are measurements in individual nearby galaxies; larger symbols represent compilations of many measurements at higher redshift, with error bars corresponding to the 1σ scatter. DCO, a galaxy with a dominant central object, which may be in part responsible for launching the gas outflow; ULIRG, ultra-luminous infrared galaxy; LBA, Lyman break analogue (local galaxies that have high star-formation rates); NGC, New Galactic Catalogue (identifying nearby galaxies). UV, Na I and Mg II refer to the measurement methods used, with UV referring to ultraviolet measurements and Na I and Mg II referring to specific resonance lines. This figure is taken from figure 14 of ref. 30, with permission.

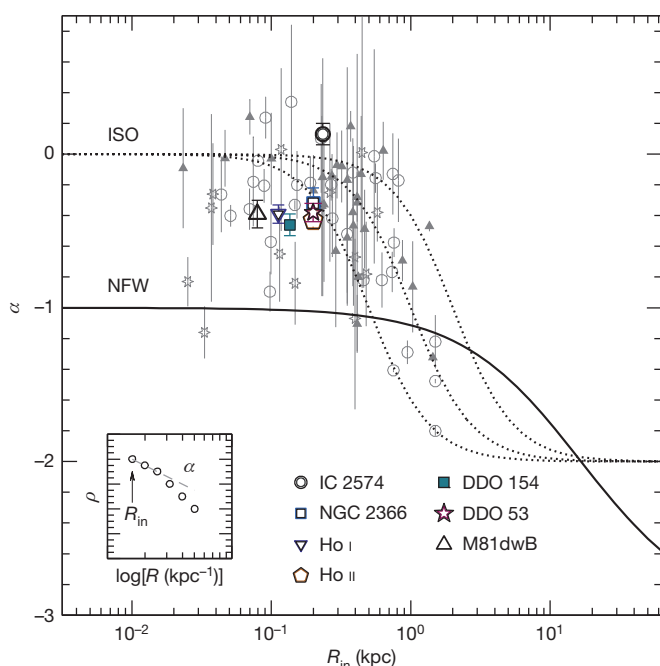


Figure 2 | Constant-density cores of dark matter in dwarf galaxies. This graph⁵⁰ shows the observed innermost dark matter density power law (an estimate of α , where the dark matter density ρ_{DM} and radius r are locally related by $\rho_{\text{DM}}(r) \propto r^\alpha$) for field dwarf galaxies, plotted at the innermost point where a robust determination has been achieved. Where the slope α can be measured within one kiloparsec, it is typically much shallower ($\alpha > -1$) than the simulated NFW result⁶. The 1σ error bars show uncertainties rising chiefly from modelling non-circular motions of the gas. Individual galaxy names are as shown. This figure is taken from figure 8 of ref. 50, with permission.

dark matter and the physics of galaxy formation and have accordingly received significant attention⁵⁵.

We discussed above how field dwarfs have been fundamental in revealing the apparent over-concentration of dark matter at the centre of halos. Satellite dwarfs, with an order-of-magnitude fewer stars, are potentially powerful probes of the dark matter distribution at the smallest scales⁵⁶. Various techniques hint at the existence of cores, rather than cusps, in the brightest dwarf spheroidals^{57–59}. However, because galaxy satellites do not possess atomic hydrogen disks and deviate from spherical symmetry, inferring the mass distribution of their dark matter halos involves different, less well-developed techniques than those applied to field galaxies. Rather than estimate the entire mass distribution, it is simpler to measure total mass inside the half-light radius (which typically lies at a few hundred parsecs)⁶⁰. Compared to the most massive satellites in cold dark matter it is widely believed that there is too little mass in each real dwarf spheroidal, a problem which is referred to as the objects being “too big to fail”⁶¹, although the effect of tidal forces and stripping^{62,63} complicate the interpretation. At present, the properties and abundance of isolated, small field galaxies provide stronger constraints on models of feedback and alternative dark matter models^{64,65}.

High-mass galaxies

Field dwarfs typically fall into the category of ‘low-surface-brightness’ galaxies, defined by their extended diffuse stellar and gaseous disks. The uncertainties (discussed above) in recovering dark matter distributions in these objects are mitigated by the relatively small baryonic contribution to the potential at the time they are observed. A fraction of more massive galaxies (with rotational peak velocities larger than 100 km s^{-1}) also have these characteristics. Analyses of such galaxies⁶⁶ again point to relatively flat central dark matter profiles. This is a significant finding because it shows that cores can be formed in halos with estimated stellar masses up to $5 \times 10^9 M_\odot$, where M_\odot is the mass of the Sun.

The inner distribution of dark matter in galaxies with more conventional, massive disks (similar to our own Milky Way, for instance) is unfortunately harder to ascertain because the gravitational potential is more strongly dominated by baryons¹⁵, so that uncertainties in the age, metallicity and hence light-to-mass conversion ratios of stellar populations dominate. However, many attempts have pointed to smaller central dark matter densities than theoretically expected⁶⁷, in line with the low-surface-brightness results. Some observations point to well-defined scaling laws that link the dark matter and baryon components, with dark matter and baryons following similar profiles⁶⁸. The significance of this relation is still very poorly understood, but it may point to a tight coupling between baryons and dark matter at galactic scales⁶⁹. More indirect constraints on the central dark matter densities in luminous galaxies arise from the existence of stellar bars⁷⁰ (that is, strong non-axisymmetric features in the distribution of stars in the galactic disk), which, over cosmological timescales, seems dynamically incompatible with the presence of cuspy dark matter halos⁷¹.

The largest bound systems in the Universe, galaxy clusters, have also attracted much interest in recent years as techniques to extract their density profile from lensing and kinematical data have developed. Subtracting the baryonic contribution to determine the presence of dark matter cusps or cores remains difficult, largely because of the unknown mass-to-light ratio of stars^{72,73}. As techniques mature further and sample sizes increase, this area will become increasingly important.

Gravitational disturbances to dark matter

We have outlined above the observational evidence pointing towards systematic departures of the distribution of dark matter from the original expectations of the cold dark matter paradigm. It has been widely suggested that this discrepancy could be addressed by gravitational interactions (the only way baryons and cold dark matter can interact) that transfer energy from the baryon component to the diffuse dark matter⁷⁴. If sufficient energy can be given gravitationally to dark matter particles in the centre of the halo, they will then migrate outwards, reducing the central density (note that this process will also apply to the stellar component^{75–78}). Energy can be transferred between these two components in two distinct ways: from the kinetic energy of incoming material or from baryonic processes linked to feedback within the galaxy. We will tackle these possibilities in turn.

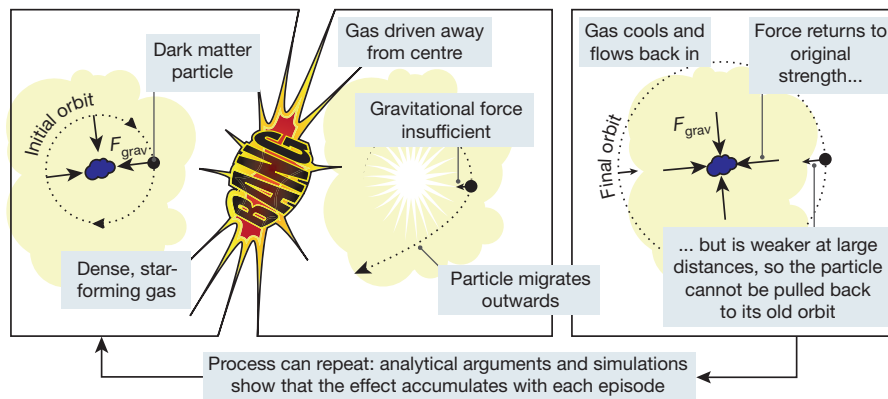
As dense clumps move through a diffuse dark matter background a fraction of the orbital energy of the incoming material is lost to internal energy of the diffuse halo through “dynamical friction”^{79–81} (see Box 1 for an explanation). The sinking of dense gaseous or stellar clumps can flatten the central dark matter profile over a range of scales, although significant core creation has been demonstrated only in simulations of galaxy clusters rather than at the scale of individual galaxies⁸². Dense, centrally concentrated baryons in in-falling clumps are an essential prerequisite for this process.

The second class of energy sources comes from within the galaxy itself: energy liberated from stellar populations can be large compared to the binding energy of the galaxies⁴³. Early work suggested that removing most of the baryons in a rapid, dramatic starburst event could overcompensate for the previous adiabatic contraction, leading to the desired effect of reducing the central dark matter density⁷⁴. Subsequent studies examined the feasibility of this mechanism in more detail^{83–85}, showing in particular that repeated outflow episodes interspersed by re-accretion had a cumulative effect on the dark matter⁸⁵.

However, these early investigations were limited by the unknown behaviour of gas in dwarf galaxies over cosmic time, and the lack of any clear analytic framework for understanding the apparently irreversible response of the dark matter. The outflows required seemed to redistribute a larger mass of gas than permitted by observations⁸³. It was unclear even to what extent the available energy in stellar populations couples to the gas through heating and radiation pressure; consequently the idea of energy transfer from baryons to the dark matter was not widely accepted at this stage.

BOX 1

Gas affects dark matter through gravity



Box 1 Figure How repeated starbursts inside galaxies cause dark matter to migrate outwards, generating a near-constant-density 'core'.

Until recently, dark matter was widely thought to contract slightly in response to galaxy formation²⁷. But the 'adiabatic' calculations giving this result assume that gas moves in gradual steady flows. If it instead arrives in discrete dense clumps, a gravitationally induced wake of dark matter pulls back on the clump. This gives a net transfer of energy from gas to dark matter, a transformative effect known as dynamical friction^{79,80}. The assumptions underlying adiabatic modelling can also fail owing to outflows if these evacuate gas at speeds significantly exceeding the local circular velocity⁷⁴. If the removal proceeds sufficiently quickly, net energy is again transferred into the dark matter^{76,85}. Moreover, this transfer is irreversible in the sense that re-accreting the lost gas does not lead to a compensating energy loss⁷⁶ (see Box 1 Figure). These results hold even if the gas never leaves the galaxy but is simply moved in bulk internally^{76,86,87}. The reason for this is as follows⁷⁶. Consider a dark matter particle that orbits close to the centre of the halo, where the gas is dense. If the gas is locally removed on a short timescale, the gravitational centripetal force holding the dark matter in its orbit is suddenly reduced in magnitude. The dark matter particle responds by flying outwards. Even if the gas later cools and settles back, the dark matter particle does not return to its original orbit. This is because the strength of gravity naturally declines with distance; once the particle has flown out to large distances, it is not possible to pull it back again. Repeating the process has a cumulative effect, which allows a significant transformation to be accomplished by recycling a small amount of gas instead of expelling an unfeasibly large amount of gas in one episode.

Other authors^{86,87} showed that gas remaining fully within the system can still be effective in removing cusps when coupled to an energy source such as stellar feedback. For instance, supernovae driving gas on timescales close to the local orbital period were identified as a mechanism by which to transfer energy to dark matter particles⁸⁷. In this case, the cusps were destroyed in an energetically consistent manner without requiring any unrealistically dramatic outflows. By 2008 advances in numerical resolution and understanding of how gas cools before forming stars allowed for realistic treatments of the relevant hydrodynamics (Box 2). Simulations at high redshift⁸⁸ showed that dark matter could indeed be expelled self-consistently from the central regions of small protogalactic objects. This work provided the first proof-of-concept in a cosmological setting, but did not make predictions of observable objects (dwarfs are so faint that they are only observable in the nearby, redshift-zero Universe).

As it became possible to resolve star-forming regions⁸⁹ throughout the assembly of a dwarf galaxy from the young universe to the present day, for the first time simulations formed galaxies with stellar, gas and dark matter distributions consistent with observational bounds^{44,90,91}. Multiple short, locally concentrated bursts of star formation were the key new phenomenon enabling modification of the dark matter distribution: by temporarily evacuating gas from the central kiloparsec of the galaxy these cause dark matter to migrate irreversibly outwards⁷⁶; see Box 1. The actual process in play thus combines characteristics of the multiple-epoch outflow picture⁸⁵ and the internal-motions picture⁸⁷. It does not require fine-tuning of the gas velocity or dramatic evacuation of the gas from anything but the innermost region. The key requirement is that the gas must exit the centre of the galaxy at speeds greater than the local circular velocity.

Analytical modelling of multiple, impulsive changes to the gravitational potential gives considerable insight into how these changes arise

and why they are irreversible⁷⁶. This allows for an accumulation of effects as the process repeats in several gas outflow events. In a single event the total gas mass in the galaxy limits the effect of outflows⁸³ but when the same gas is recycled and used in multiple events the only practical limitation is the total energy liberated from stellar populations and black holes (see below). The model of core creation through repeated outflows draws strong support from both analytical arguments⁷⁶ and from simulations using different numerical techniques⁷⁸. Observationally, dwarf galaxies, in which the evidence for cores is strongest, are observed to be gas-rich and show evidence for repeated small bursts and prolonged star-formation histories⁹². This supports a picture in which the effect on the dark matter builds up over several billion years^{76,85}, during which gas is being cycled in repeating outflow and cooling episodes.

Core sizes depend on galaxy mass

A key part of confirming which mechanisms are responsible for flattened dark matter profiles is to predict and understand in detail how the processes affect systems of differing size. Building on the impulsive picture⁷⁶, full numerical simulations⁹⁰ and analytical arguments⁹³ have all pointed to a transition between core creation and persistent cusps below a critical stellar mass. This dividing line probably lies between $10^6 M_{\odot}$ and $10^7 M_{\odot}$ (assuming that most of the energy available from supernovae is transferred to the dark matter). For less massive stellar systems, the direct effects of stellar feedback on the dark matter should be minor on energetic grounds alone⁹³, as star formation becomes less efficient; see Fig. 3. The energetic argument shows that the possible cores from supernova feedback would be undetectably small for stellar masses significantly below $10^6 M_{\odot}$.

For stellar masses exceeding $10^7 M_{\odot}$, it is clear that energy from star-formation processes is available to alter the central regions of the dark

BOX 2

How to generate outflows

Outflows are probably generated by young stars inside galaxies. Computer simulations of the formation of galaxies would therefore ideally resolve cosmological large-scale structure (over tens of megaparsecs) down to the scale of individual stars (at least 10^{14} times smaller). This is, and seems certain to remain, unfeasible. The line of attack is instead to mimic the effects of stars without actually resolving them individually. Because star formation is the conclusion of runaway gas cooling and collapse, a typical computational approach is to form stars when gas satisfies certain averaged conditions, and in particular when it reaches a threshold density. As resolution slowly improves in simulations, smaller regions and larger densities can be self-consistently resolved⁸⁹. Until the mid-2000s, a typical threshold density was set at $0.1 m_{\text{H}} \text{cm}^{-3}$, where m_{H} is the mass of a hydrogen atom. This corresponds to the mean density of galactic neutral atomic gas, so stars form throughout the disk of a simulated galaxy. Energy output from stars in the diffuse medium results in a gentle heating of the entire galaxy, slowing the process of further star formation. However, if one can achieve sufficient resolution (and implement the more complicated cooling physics required^{16,38,110}) to push to $10 m_{\text{H}} \text{cm}^{-3}$ or $100 m_{\text{H}} \text{cm}^{-3}$, then a qualitatively different behaviour results. This is the density that corresponds to molecular cloud formation in our Galaxy, known to be the precursor of star formation. Instead of forming stars in a diffuse way through the entire disk, now stars form efficiently in small, isolated regions^{23,44}, which is considerably more realistic. When energy from the resulting stellar populations is dumped into the gas, the cloud heats to much higher temperatures than diffuse star formation achieves. It is likely that intense radiation pressure is also a significant factor³⁴. In any case, the gas is overpressurized by a factor of at least a hundred compared to its surroundings and expands rapidly. The combination of high initial density and explosive decompression is suitable for launching galactic-scale outflows; it is also what allows an efficient coupling of the available energy to dark matter (Box 1).

matter halo through sufficiently rapid galactic fountains or outflows⁹⁰, but few simulations of luminous galaxies reach the resolution necessary to study the formation of cores. Some high-resolution simulations of Milky Way analogues have been reported to form dark matter cores on scales of a kiloparsec or larger^{94,95}. On the other hand it has been reported that cores shrink with respect to the halo scale radius⁹⁶ for total masses exceeding $10^{11} M_{\odot}$ (the mass of the Milky Way is about $10^{12} M_{\odot}$). These statements may be reconcilable; further higher-resolution work is required for progress in our understanding. As masses continue to increase to the cluster scale (see the ‘High-mass galaxies’ section above), further processes become interesting. For instance, numerical work has shown that accretion onto the central black hole, if proceeding in repeated, highly energetic bursts, replicates the effect of supernovae on dwarf galaxies⁹⁷.

Modifying dark matter

We have established that there are many processes that can modify the dark matter distribution in the centre of galaxies, even if the dark matter is cold and collisionless (that is, interacts only through gravity)—a ‘minimal’ scenario motivated by supersymmetric weakly interacting massive particles. However, the observational controversies detailed in the ‘Evidence for a cusp–core discrepancy’ section above have prompted considerable interest in non-minimal dark matter models. By changing the properties of the dark matter candidate particle, the predictions for the distribution within halos are altered; potentially, therefore, galaxies and galaxy clusters become an important probe of particle physics.

For instance, the warm dark matter models (WDM) invoke a candidate particle with non-negligible residual streaming motions after decoupling

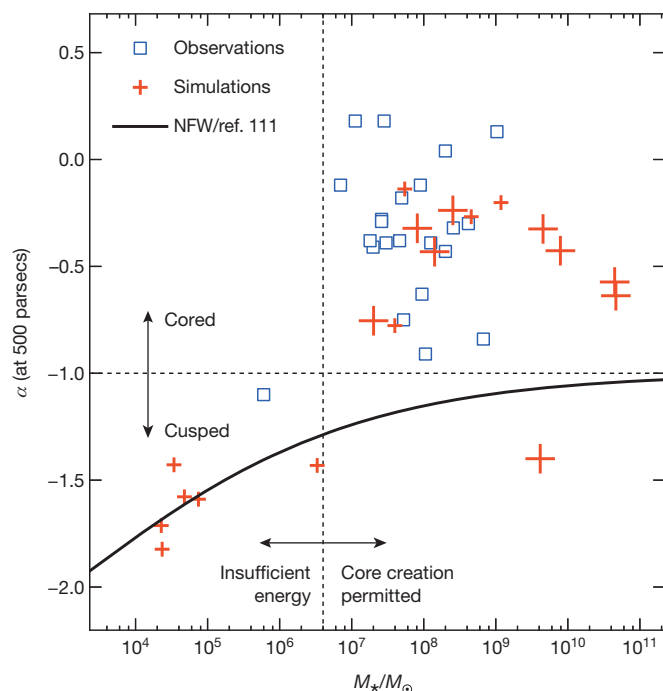


Figure 3 | Dark matter cores are only generated in sufficiently bright galaxies. Here we have plotted the power-law index α of the dark matter density (as in Fig. 2, but here measured at radius 500 parsecs) against the mass of stars formed, M_* (updated from ref. 90). The expected slopes from pure dark matter calculations are approximated by the solid line (using the scaling relations from ref. 111), whereas hydrodynamic simulations at high mass have shallower slopes, indicated by the crosses. Large crosses show halos resolved with more than 500,000 simulated dark matter particles. Smaller crosses have fewer particles, but always more than 50,000. When less than about $10^{6.5} M_{\odot}$ of gas has formed into stars, there is insufficient energy available to flatten the cusp⁹³. The box symbols show data from the THINGS survey⁵⁰ of field dwarf galaxies. Additional observational data at stellar masses lower than $10^6 M_{\odot}$ would be highly valuable. This figure is updated from figure 1 of ref. 90.

(such as a sterile neutrino), suppressing the formation of small-scale structure⁹⁸ and delaying the collapse of dwarf-sized halos and their associated star formation to slightly later epochs⁹⁹. However, these models do not produce cores on observationally relevant scales¹⁰⁰ and are currently strongly constrained by the clustering of the neutral gas in the cosmic web¹⁰¹.

Another major class, self-interacting dark matter (SIDM)¹⁰², refers to particle physics scenarios with significant ‘dark sector’ interactions. SIDM behaves more like a collisional fluid, preventing the central high-density cusp from forming and making the central regions more spherical¹⁰³. Unlike in the WDM case, the number density of dark matter halos remains relatively unchanged even at the smallest scales¹⁰⁴. The diversity of theoretical models, however, gives significant freedom in the choice of the cross-section and its possible dependence on particle velocity¹⁰⁵. This makes it difficult to establish a single baseline SIDM scenario.

The majority of work on non-minimal dark matter falls into the WDM or SIDM categories. However, modifications to the dark matter profile can also be achieved through other processes. For instance, particle–particle annihilations¹⁰⁶ can reduce central densities directly, provided the physics is tuned to prevent rapid annihilation in the early universe. Alternatively, if dark matter decays over long timescales to slightly lighter daughter particles, the lost mass provides a source of kinetic energy for expanding the centre of dark matter halos¹⁰⁷. Another relevant possibility is that the dark matter is not formed from particles at all. In the case of an ultralight scalar field, for instance, the Compton wavelength becomes larger than the supposed interparticle separation; accordingly the field behaves as a Bose–Einstein condensate¹⁰⁸ rather than as individual particles, preventing the central cusps from forming.

The discussion above is not intended to be exhaustive, but to indicate that there is no shortage of particle physics mechanisms that address the density of dark matter in galactic cores. However, unlike baryonic mechanisms, these do not simultaneously address the gas ‘angular momentum catastrophe’ discussed above. The angular momentum of gas arises from the gravitational collapse process, and so—in the absence of strong baryonic feedback—is fundamentally tied to the nature of large-scale structure rather than to details of the collapsed halos into which it flows. Given constraints from the distribution of intergalactic hydrogen, as seen through the Lyman- α absorption in quasar spectra, permitted small-scale modifications to the power spectrum do not lead to structures with increased angular momentum¹⁰⁰ and so cannot resolve the gas ‘angular momentum catastrophe’ discussed above.

This said, it would be foolhardy to proceed on the basis that cold dark matter must describe the correct fundamental particle physics. In particular, the effects of baryons may amplify or change the signatures of alternative particle models. The dwarf spheroidals teach us that different transformative mechanisms interact in surprising, nonlinear ways⁶², motivating detailed study of the galaxies formed in fully hydrodynamical simulations with exotic dark matter models¹⁰⁹, a field that remains in its infancy relative to that of galaxy formation based on cold dark matter.

Priorities for future work

There are now plenty of known processes that can modify the distribution of dark matter, either by supposing non-minimal particle physics or by harnessing the power released by stars and black holes. The priority is to establish which of these are in play in the real Universe. Because the baryonic processes simultaneously modify a number of observational diagnostics (outflows, dark matter cores, stellar morphology and star-formation regulation), they weave into a coherent, testable framework. Additionally, to alleviate degeneracies between particle-physics-induced and outflow-induced modifications to cold dark matter, one can identify regimes in which only one or the other is active. This points towards the future value of careful studies probing scalings of cores from stellar masses below $10^7 M_\odot$ (where the energy available to create cores is so limited that baryonic effects are tightly constrained) to above $10^{13} M_\odot$ (where a variety of processes are feasible).

It remains a possibility that tensions between observation and theory will begin to point unambiguously at exotic particle physics. Ultimately, we expect that a concerted effort from theorists and observers can achieve the goal of pointing to unique predictions of non-minimal dark matter models. Of particular interest in the coming years will be: (1) an improved understanding of the dark matter in dwarf spheroidals and faint field galaxies; if cores persist at the faintest end, it is a generic conclusion that baryonic physics cannot account for them^{90,93}; (2) the study of the stellar population ages and, separately, metallicity distributions of these objects to determine as far as possible whether the required bursty star-formation histories are consistent propositions⁹²; (3) better predictions of the scalings of cores in massive galaxies and clusters for different scenarios; (4) observations that constrain the star-formation histories of dwarfs and the behaviour of gas at high redshift, especially through absorption line studies, which are sensitive to internal kinematics and outflows¹⁰¹; and (5) a renewed effort to understand how non-minimal dark matter scenarios (such as WDM or SIDM) interact with the revised, more complex baryonic physics of galaxy formation¹⁰⁹.

Received 1 October; accepted 4 December 2013.

1. The Planck Collaboration *et al.* Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/1303.5076>.
2. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311**, 517–525 (1984).
3. Percival, W. J. *et al.* The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *Mon. Not. R. Astron. Soc.* **327**, 1297–1306 (2001).
4. White, S. D. M. & Rees, M. J. Core condensation in heavy halos—a two-stage theory for galaxy formation and clustering. *Mon. Not. R. Astron. Soc.* **183**, 341–358 (1978).

This paper established the paradigm in which visible galaxies form in the gravitational potential carved out by dark matter halos.

5. White, S. D. M., Frenk, C. S. & Davis, M. Clustering in a neutrino-dominated universe. *Astrophys. J.* **274**, L1–L5 (1983).
6. Navarro, J. F., Frenk, C. S. & White, S. D. M. The structure of cold dark matter halos. *Astrophys. J.* **462**, 563–575 (1996).
- This work emphasized the universality of the behaviour of simulated dark matter structures, including a central ‘cusp’, irrespective of scale.**
7. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
8. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where are the missing galactic satellites? *Astrophys. J.* **522**, 82–92 (1999).
9. Mateo, M. L. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).
10. Dubinski, J. & Carlberg, R. G. The structure of cold dark matter halos. *Astrophys. J.* **378**, 496–503 (1991).
- This early paper clearly identified the ‘cusp’ in simulations of the behaviour of cold dark matter.**
11. Moore, B., Governato, F., Quinn, T., Stadel, J. & Lake, G. Resolving the structure of cold dark matter halos. *Astrophys. J.* **499**, L5–L8 (1998).
12. de Blok, W. J. G. *et al.* High-resolution rotation curves and galaxy mass models from THINGS. *Astron. J.* **136**, 2648–2719 (2008).
13. Springel, V. *et al.* The Aquarius Project: the subhaloes of galactic haloes. *Mon. Not. R. Astron. Soc.* **391**, 1685–1711 (2008).
14. Stadel, J. *et al.* Quantifying the heart of darkness with GHALO—a multibillion particle simulation of a galactic halo. *Mon. Not. R. Astron. Soc.* **398**, L21–L25 (2009).
15. Bell, E. F. & de Jong, R. S. Stellar mass-to-light ratios and the Tully-Fisher relation. *Astrophys. J.* **550**, 212–229 (2001).
16. Gnedin, N. Y., Tassis, K. & Kravtsov, A. V. Modeling molecular hydrogen and star formation in cosmological simulations. *Astrophys. J.* **697**, 55–67 (2009).
17. Wise, J. H. & Abel, T. ENZO+MORAY: radiation hydrodynamics adaptive mesh refinement simulations with adaptive ray tracing. *Mon. Not. R. Astron. Soc.* **414**, 3458–3491 (2011).
18. Teyssier, R. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *Astron. Astrophys.* **385**, 337–364 (2002).
19. Kereš, D., Vogelsberger, M., Sijacki, D., Springel, V. & Hernquist, L. Moving-mesh cosmology: characteristics of galaxies and haloes. *Mon. Not. R. Astron. Soc.* **425**, 2027–2048 (2012).
20. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365**, 11–28 (2006).
21. Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655 (2006).
22. Stinson, G. *et al.* Star formation and feedback in smoothed particle hydrodynamic simulations—I. Isolated galaxies. *Mon. Not. R. Astron. Soc.* **373**, 1074–1090 (2006).
23. Hopkins, P. F., Quataert, E. & Murray, N. Stellar feedback in galaxies and the origin of galaxy-scale winds. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
24. Robertson, B. *et al.* A merger-driven scenario for cosmological disk galaxy formation. *Astrophys. J.* **645**, 986–1000 (2006).
25. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
26. Dekel, A. *et al.* Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).
27. Blumenthal, G. R., Faber, S. M., Flores, R. & Primack, J. R. Contraction of dark matter galactic halos due to baryonic infall. *Astrophys. J.* **301**, 27–34 (1986).
28. Shapley, A. E., Steidel, C. C., Pettini, M. & Adelberger, K. L. Rest-frame ultraviolet spectra of z ~ 3 Lyman break galaxies. *Astrophys. J.* **588**, 65–89 (2003).
29. Weiner, B. J. *et al.* Ubiquitous outflows in DEEP2 spectra of star-forming galaxies at z = 1.4. *Astrophys. J.* **692**, 187–211 (2009).
30. Martin, C. L. *et al.* Demographics and physical properties of gas out/inflows at 0.4. *Astrophys. J.* **760**, 127 (2012).
31. Steidel, C. C. *et al.* The structure and kinematics of the circumgalactic medium from far-ultraviolet spectra of z ~ 2–3 galaxies. *Astrophys. J.* **717**, 289–322 (2010).
32. Heckman, T. M., Lehnert, M. D., Strickland, D. K. & Armus, L. Absorption-line probes of gas and dust in galactic superwinds. *Astrophys. J.* (Suppl.) **129**, 493–516 (2000).
33. Mac Low, M. & Ferrara, A. Starburst-driven mass loss from dwarf galaxies: efficiency and metal ejection. *Astrophys. J.* **513**, 142–155 (1999).
34. Murray, N., Quataert, E. & Thompson, T. A. The disruption of giant molecular clouds by radiation pressure and the efficiency of star formation in galaxies. *Astrophys. J.* **709**, 191–209 (2010).
35. van der Wel, A. *et al.* Extreme emission-line galaxies in CANDELS: broadband-selected, starbursting dwarf galaxies at z > 1. *Astrophys. J.* **742**, 111 (2011).
36. Rubin, K. H. R. *et al.* The persistence of cool galactic winds in high stellar mass galaxies between z ~ 1.4 and ~ 1. *Astrophys. J.* **719**, 1503–1525 (2010).
37. Rubin, K. H. R., Prochaska, J. X., Koo, D. C. & Phillips, A. C. The direct detection of cool, metal-enriched gas accretion onto galaxies at z ~ 0.5. *Astrophys. J.* **747**, L26 (2012).
38. Shen, S., Wadsley, J. & Stinson, G. The enrichment of the intergalactic medium with adiabatic feedback—I. Metal cooling and metal diffusion. *Mon. Not. R. Astron. Soc.* **407**, 1581–1596 (2010).
39. Dutton, A. A. On the origin of exponential galaxy discs. *Mon. Not. R. Astron. Soc.* **396**, 121–140 (2009).

40. Kormendy, J., Drory, N., Bender, R. & Cornell, M. E. Bulgeless giant galaxies challenge our picture of galaxy formation by hierarchical clustering. *Astrophys. J.* **723**, 54–80 (2010).
41. Barnes, J. & Efstathiou, G. Angular momentum from tidal torques. *Astrophys. J.* **319**, 575–600 (1987).
42. van den Bosch, F. C., Burkert, A. & Swaters, R. A. The angular momentum content of dwarf galaxies: new challenges for the theory of galaxy formation. *Mon. Not. R. Astron. Soc.* **326**, 1205–1215 (2001).
43. Binney, J., Gerhard, O. & Silk, J. The dark matter problem in disc galaxies. *Mon. Not. R. Astron. Soc.* **321**, 471–474 (2001).
44. Governato, F. *et al.* Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows. *Nature* **463**, 203–206 (2010).
This work produced, for the first time, a simulated dwarf galaxy with stellar and dark matter distribution consistent with modern day observations.
45. Brook, C. B. *et al.* Hierarchical formation of bulgeless galaxies: why outflows have low angular momentum. *Mon. Not. R. Astron. Soc.* **415**, 1051–1060 (2011).
46. Flores, R. A. & Primack, J. R. Observational and theoretical constraints on singular dark matter halos. *Astrophys. J.* **427**, L1–L4 (1994).
47. Moore, B. Evidence against dissipation-less dark matter from observations of galaxy haloes. *Nature* **370**, 629–631 (1994).
These two papers (refs 46 and 47) pointed out the great difficulty in reconciling theoretical predictions of dark matter in dwarf galaxies with observations.
48. Simon, J. D., Bolatto, A. D., Leroy, A., Blitz, L. & Gates, E. L. High-resolution measurements of the halos of four dark matter-dominated galaxies: deviations from a universal density profile. *Astrophys. J.* **621**, 757–776 (2005).
49. Swaters, R. A., Madore, B. F., van den Bosch, F. C. & Balcells, M. The central mass distribution in dwarf and low surface brightness galaxies. *Astrophys. J.* **583**, 732–751 (2003).
50. Oh, S.-H., de Blok, W. J. G., Brinks, E., Walter, F. & Kennicutt, R. C. Jr. Dark and luminous matter in THINGS dwarf galaxies. *Astron. J.* **141**, 193 (2011).
51. Oh, S.-H. *et al.* The central slope of dark matter cores in dwarf galaxies: simulations versus THINGS. *Astron. J.* **142**, 24 (2011).
52. Walter, F. *et al.* THINGS: The H I nearby galaxy survey. *Astron. J.* **136**, 2563–2647 (2008).
53. Hunter, D. A. *et al.* LITTLE THINGS. *Astron. J.* **144**, 134 (2012).
54. Mayer, L. *et al.* The metamorphosis of tidally stirred dwarf galaxies. *Astrophys. J.* **559**, 754–784 (2001).
55. Nierenberg, A. M. *et al.* Luminous satellites. II. Spatial distribution, luminosity function, and cosmic evolution. *Astrophys. J.* **752**, 99 (2012).
56. Strigari, L. E. *et al.* A common mass scale for satellite galaxies of the Milky Way. *Nature* **454**, 1096–1097 (2008).
57. Goerdt, T., Moore, B., Read, J. I., Stadel, J. & Zemp, M. Does the Fornax dwarf spheroidal have a central cusp or core? *Mon. Not. R. Astron. Soc.* **368**, 1073–1077 (2006).
58. Walker, M. G. & Peñarrubia, J. A method for measuring (slopes of) the mass profiles of dwarf spheroidal galaxies. *Astrophys. J.* **742**, 20 (2011).
59. Breddels, M. A. & Helmi, A. Model comparison of the dark matter profiles of Fornax, Sculptor, Carina and Sextans. *Astron. Astrophys.* **558**, A35 (2013).
60. Wolf, J. *et al.* Accurate masses for dispersion-supported galaxies. *Mon. Not. R. Astron. Soc.* **406**, 1220–1237 (2010).
61. Boylan-Kolchin, M., Bullock, J. S. & Kaplinghat, M. The Milky Way's bright satellites as an apparent failure of Λ CDM. *Mon. Not. R. Astron. Soc.* **422**, 1203–1218 (2012).
62. Zolotov, A. *et al.* Baryons matter: why luminous satellite galaxies have reduced central masses. *Astrophys. J.* **761**, 71 (2012).
63. Arraki, K. S., Klypin, A., More, S. & Trujillo-Gomez, S. Effects of baryon removal on the structure of dwarf spheroidal galaxies. *Mon. Not. R. Astron. Soc.* <http://dx.doi.org/10.1093/mnras/stt2279> (in the press); preprint at <http://arxiv.org/abs/1212.6651>.
64. Papastergis, E., Martin, A. M., Giovanelli, R. & Haynes, M. P. The velocity width function of galaxies from the 40% ALFALFA survey: shedding light on the cold dark matter overabundance problem. *Astrophys. J.* **739**, 38 (2011).
65. Ferrero, I., Abadi, M. G., Navarro, J. F., Sales, L. V. & Gurovich, S. The dark matter haloes of dwarf galaxies: a challenge for the Λ cold dark matter paradigm? *Mon. Not. R. Astron. Soc.* **425**, 2817–2823 (2012).
66. Kuzio de Naray, R. & Spekkens, K. Do baryons alter the halos of low surface brightness galaxies? *Astrophys. J.* **741**, L29 (2011).
67. McGaugh, S. S., de Blok, W. J. G., Schombert, J. M., Kuzio de Naray, R. & Kim, J. H. The rotation velocity attributable to dark matter at intermediate radii in disk galaxies. *Astrophys. J.* **659**, 149–161 (2007).
68. Gentile, G., Famaey, B., Zhao, H. & Salucci, P. Universality of galactic surface densities within one dark halo scale-length. *Nature* **461**, 627–628 (2009).
69. Swaters, R. A., Sancisi, R., van Albada, T. S. & van der Hulst, J. M. Are dwarf galaxies dominated by dark matter? *Astrophys. J.* **729**, 118 (2011).
70. Pérez, I., Aguerri, J. A. L. & Méndez-Abreu, J. Bar pattern speed evolution over the last 7 Gyr. *Astron. Astrophys.* **540**, A103 (2012).
71. Debattista, V. P. & Sellwood, J. A. Dynamical friction and the distribution of dark matter in barred galaxies. *Astrophys. J.* **493**, L5 (1998).
72. Newman, A. B., Treu, T., Ellis, R. S. & Sand, D. J. The density profiles of massive, relaxed galaxy clusters. II. Separating luminous and dark matter in cluster cores. *Astrophys. J.* **765**, 25 (2013).
73. Dutton, A. A., Macciò, A. V., Mendel, J. T. & Simard, L. Universal IMF versus dark halo response in early-type galaxies: breaking the degeneracy with the fundamental plane. *Mon. Not. R. Astron. Soc.* **432**, 2496–2511 (2013).
74. Navarro, J. F., Eke, V. R. & Frenk, C. S. The cores of dwarf galaxy haloes. *Mon. Not. R. Astron. Soc.* **283**, L72–L78 (1996).
This work suggested that sufficiently violent gas disruption could change the behaviour of dark matter.
75. Read, J. I. & Gilmore, G. Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles. *Mon. Not. R. Astron. Soc.* **356**, 107–124 (2005).
76. Pontzen, A. & Governato, F. How supernova feedback turns dark matter cusps into cores. *Mon. Not. R. Astron. Soc.* **421**, 3464–3471 (2012).
By using analytic modelling, we were able to link simulated dark matter cores⁴⁴ to the physical behaviour of gas when star formation is concentrated and bursty.
77. Maxwell, A. J., Wadsley, J., Couchman, H. M. P. & Mashchenko, S. Building the stellar halo through feedback in dwarf galaxies. *Astrophys. J.* **755**, L35 (2012).
78. Teyssier, R., Pontzen, A., Dubois, Y. & Read, J. I. Cusp-core transformations in dwarf galaxies: observational predictions. *Mon. Not. R. Astron. Soc.* **429**, 3068–3078 (2013).
79. White, S. D. M. Dynamical friction in spherical clusters. *Mon. Not. R. Astron. Soc.* **174**, 19–28 (1976).
80. El-Zant, A., Shlosman, I. & Hoffman, Y. Dark halos: the flattening of the density cusp by dynamical friction. *Astrophys. J.* **560**, 636–643 (2001).
81. Tonini, C., Lapi, A. & Salucci, P. Angular momentum transfer in dark matter halos: erasing the cusp. *Astrophys. J.* **649**, 591–598 (2006).
82. El-Zant, A. A., Hoffman, Y., Primack, J., Combes, F. & Shlosman, I. Flat-cored dark matter in cuspy clusters of galaxies. *Astrophys. J.* **607**, L75–L78 (2004).
83. Gnedin, O. Y. & Zhao, H. Maximum feedback and dark matter profiles of dwarf galaxies. *Mon. Not. R. Astron. Soc.* **333**, 299–306 (2002).
84. Mo, H. J. & Mao, S. Galaxy formation in pre-processed dark haloes. *Mon. Not. R. Astron. Soc.* **353**, 829–840 (2004).
85. Read, J. I. & Gilmore, G. Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles. *Mon. Not. R. Astron. Soc.* **356**, 107–124 (2005).
86. Weinberg, M. D. & Katz, N. Bar-driven dark halo evolution: a resolution of the cusp-core controversy. *Astrophys. J.* **580**, 627–633 (2002).
87. Mashchenko, S., Couchman, H. M. P. & Wadsley, J. The removal of cusps from galaxy centres by stellar feedback in the early Universe. *Nature* **442**, 539–542 (2006).
88. Mashchenko, S., Wadsley, J. & Couchman, H. M. P. Stellar feedback in dwarf galaxy formation. *Science* **319**, 174–177 (2008).
These authors presented the first self-consistent simulations in which galaxies lost their central cusps due to supernova feedback.
89. Saitoh, T. R. *et al.* Toward first-principle simulations of galaxy formation: I. How should we choose star-formation criteria in high-resolution simulations of disk galaxies? *Pub. Astron. Soc. Jpn* **60**, 667–681 (2008).
90. Governato, F. *et al.* Cuspy no more: how outflows affect the central dark matter and baryon distribution in Λ cold dark matter galaxies. *Mon. Not. R. Astron. Soc.* **422**, 1231–1240 (2012).
91. Munshi, F. *et al.* Reproducing the stellar mass/halo mass relation in simulated Λ CDM galaxies: theory vs observational estimates. *Astrophys. J.* **766**, 56 (2013).
92. McQuinn, K. B. W. *et al.* The nature of starbursts. I. The star formation histories of eighteen nearby starburst dwarf galaxies. *Astrophys. J.* **721**, 297–317 (2010).
93. Peñarrubia, J., Pontzen, A., Walker, M. G. & Koposov, S. E. The coupling between the core/cusp and missing satellite problems. *Astrophys. J.* **759**, L42 (2012).
94. Macciò, A. V. *et al.* Halo expansion in cosmological hydro simulations: toward a baryonic solution of the cusp/core problem in massive spirals. *Astrophys. J.* **744**, L9 (2012).
95. Kuhlen, M., Guedes, J., Pillepich, A., Madau, P. & Mayer, L. An off-center density peak in the Milky Way's dark matter halo? *Astrophys. J.* **765**, 10 (2013).
96. Di Cintio, A. *et al.* The dependence of dark matter profiles on the stellar to halo mass ratio: a prediction for cusps vs cores. *Mon. Not. R. Astron. Soc.* **437**, 415–423 (2014).
97. Martizzi, D., Teyssier, R., Moore, B. & Wentz, T. The effects of baryon physics, black holes and active galactic nucleus feedback on the mass distribution in clusters of galaxies. *Mon. Not. R. Astron. Soc.* **422**, 3081–3091 (2012).
98. Dalcanton, J. J. & Hogan, C. J. Halo cores and phase-space densities: observational constraints on dark matter physics and structure formation. *Astrophys. J.* **561**, 35–45 (2001).
99. Menci, N., Fiore, F. & Lamastra, A. Galaxy formation in warm dark matter cosmology. *Mon. Not. R. Astron. Soc.* **421**, 2384–2394 (2012).
100. Knebe, A., Devriendt, J. E. G., Mahmood, A. & Silk, J. Merger histories in warm dark matter structure formation scenarios. *Mon. Not. R. Astron. Soc.* **329**, 813–828 (2002).
101. Viel, M., Becker, G. D., Bolton, J. S. & Haehnelt, M. G. Warm dark matter as a solution to the small scale crisis: new constraints from high redshift Lyman- α forest data. *Phys. Rev. D* **88**, 043502 (2013).
102. Spergel, D. N. & Steinhardt, P. J. Observational evidence for self-interacting cold dark matter. *Phys. Rev. Lett.* **84**, 3760–3763 (2000).
103. Peter, A. H. G., Rocha, M., Bullock, J. S. & Kaplinghat, M. Cosmological simulations with self-interacting dark matter. II: Halo shapes vs. observations. *Mon. Not. R. Astron. Soc.* **430**, 105–120 (2013).
104. Zavala, J., Vogelsberger, M. & Walker, M. G. Constraining self-interacting dark matter with the Milky Way's dwarf spheroidals. *Mon. Not. R. Astron. Soc.* **431**, L20–L24 (2013).

105. Tulin, S., Yu, H.-B. & Zurek, K. M. Resonant dark forces and small scale structure. *Phys. Rev. Lett.* **110**, 111301 (2013).
106. Kaplinghat, M., Knox, L. & Turner, M. S. Annihilating cold dark matter. *Phys. Rev. Lett.* **85**, 3335–3338 (2000).
107. Peter, A. H. G., Moody, C. E. & Kamionkowski, M. Dark-matter decays and self-gravitating halos. *Phys. Rev. D* **81**, 103501 (2010).
108. Sin, S.-J. Late-time phase transition and the galactic halo as a Bose liquid. *Phys. Rev. D* **50**, 3650–3654 (1994).
109. Herpich, J. *et al.* MaGICC-WDM: the effects of warm dark matter in hydrodynamical simulations of disc galaxy formation. *Mon. Not. R. Astron. Soc.* **437**, 293–304 (2014).
110. Christensen, C. *et al.* Implementing molecular hydrogen in hydrodynamic simulations of galaxy formation. *Mon. Not. R. Astron. Soc.* **425**, 3058–3076 (2012).
111. Macciò, A. *et al.* Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment. *Mon. Not. R. Astron. Soc.* **378**, 55–71 (2007).

Acknowledgements We thank S.-H. Oh, S. White, M. Pettini, C. Martin, M. Walker, J. Peñarrubia, A. Brooks, T. Treu, R. Ellis, J. Wadsley and L. Randall for discussions and comments on an early draft. A.P. acknowledges support from the Oxford Martin School and Royal Society. F.G. acknowledges support from HST GO-1125 and NSF AST-0908499.

Author Contributions A.P. and F.G. jointly wrote the Review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to A.P. (a.pontzen@ucl.ac.uk).

De novo mutations in schizophrenia implicate synaptic networks

Menachem Fromer^{1,2}, Andrew J. Pocklington³, David H. Kavanagh³, Hywel J. Williams³, Sarah Dwyer³, Padhraig Gormley^{4,5}, Lyudmila Georgieva³, Elliott Rees³, Priit Palta^{4,6,7}, Douglas M. Ruderfer^{1,3}, Noa Carrera³, Isla Humphreys³, Jessica S. Johnson¹, Panos Roussos¹, Douglas D. Barker², Eric Banks⁵, Vihra Milanova⁸, Seth G. Grant⁹, Eilis Hannon³, Samuel A. Rose², Kimberly Chambert², Milind Mahajan¹, Edward M. Scolnick², Jennifer L. Moran², George Kirov³, Aarno Palotie^{4,5,7}, Steven A. McCarroll^{2,5,10}, Peter Holmans³, Pamela Sklar^{1,11}, Michael J. Owen³, Shaun M. Purcell^{1,2,12} & Michael C. O'Donovan³

Inherited alleles account for most of the genetic risk for schizophrenia. However, new (*de novo*) mutations, in the form of large chromosomal copy number changes, occur in a small fraction of cases and disproportionately disrupt genes encoding postsynaptic proteins. Here we show that small *de novo* mutations, affecting one or a few nucleotides, are overrepresented among glutamatergic postsynaptic proteins comprising activity-regulated cytoskeleton-associated protein (ARC) and N-methyl-D-aspartate receptor (NMDAR) complexes. Mutations are additionally enriched in proteins that interact with these complexes to modulate synaptic strength, namely proteins regulating actin filament dynamics and those whose messenger RNAs are targets of fragile X mental retardation protein (FMRP). Genes affected by mutations in schizophrenia overlap those mutated in autism and intellectual disability, as do mutation-enriched synaptic pathways. Aligning our findings with a parallel case-control study, we demonstrate reproducible insights into aetiological mechanisms for schizophrenia and reveal pathophysiology shared with other neurodevelopmental disorders.

Schizophrenia is a disorder whose pathophysiology is largely unknown. It has a lifetime risk of about 1%, is frequently chronic and socially disabling, and is associated with an average reduction in lifespan of about 25 years. High heritability points to a major role for transmitted genetic variants¹. However, it is also associated with a marked reduction in fecundity², leading to the hypothesis that alleles with large effects on risk might often occur *de novo* (mutations present in affected individual but not in either parent) to balance their elimination from the population by selection³.

Of the known risk alleles for schizophrenia, the only ones definitively shown to confer considerable increments in risk are rare chromosomal copy number variants (CNVs)^{1,4}, which involve deletion or duplication of thousands of bases of DNA. As predicted by the association of schizophrenia with decreased fecundity, these CNVs often occur *de novo* in the small proportion of cases in which they are found⁵. Exome sequencing technology now allows systematic scans of genes for *de novo* mutations at single-base rather than kilobase resolution. This approach has already implicated *de novo* loss-of-function mutations in disorders in which, as in schizophrenia, *de novo* CNVs have a role, including autism spectrum disorder (ASD)^{6–9} and intellectual disability^{10,11}. In schizophrenia, the results from exome sequencing^{12–14} do not yet support definitive conclusions, probably owing to limited sample sizes.

We report the largest exome sequencing study of *de novo* mutations in schizophrenia to date, using genomic (blood) DNA from 623 schizophrenia trios. The primary aims were fourfold (Table 1a–d). The first two aims were to establish a general case for the relevance of *de novo* mutations in schizophrenia by determining whether *de novo* mutations

affecting protein sequences occur in schizophrenia at higher than expected rates (Table 1a) or are enriched among sets of genes implicated in the disorder through other approaches (Table 1b). The remaining two aims, the main motivation for the study, were to determine whether *de novo* mutations implicate specific pathogenic biological processes in schizophrenia (Table 1c) and to investigate the relationship between schizophrenia and other neurodevelopmental disorders (Table 1d). To test for reproducibility, and ensure robustness of the findings to study design, we shared our findings with an independent case-control exome sequencing study¹⁵.

De novo mutation rates

We generated sequence data for a median of 93% of targeted exome bases at a depth of >10 reads, from which we generated putative *de novo* calls (Extended Data Figs 1 and 2; Supplementary Information). Using Sanger sequencing, we validated 637 *de novo* coding or canonical splice site variants (Supplementary Table 1) in 617 probands (6 trios were excluded after quality control), a rate of 1.032 mutations per trio. These comprised 482 nonsynonymous mutations, of which 64 were loss-of-function (nonsense, splice and frameshift). The remaining 155 mutations were silent and were therefore excluded from enrichment analyses.

The exome point-mutation rate in schizophrenia was, adjusting for target coverage, 1.61×10^{-8} per base per generation, compatible with the population expectation of 1.64×10^{-8} (Supplementary Information). The mutation rate (corrected for experimental confounders, Supplementary Information) was associated with increasing paternal ($P = 0.005$) and maternal ($P = 0.0003$) age at proband birth. Given the high correlation

¹Division of Psychiatric Genomics in the Department of Psychiatry, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff CF24 4HQ, UK. ⁴Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ⁵Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁶Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia. ⁷Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00290 Helsinki, Finland. ⁸Department of Psychiatry, Medical University, Sofia 1431, Bulgaria. ⁹Centre for Neuroregeneration, University of Edinburgh, Edinburgh EH16 4SB, UK. ¹⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ¹²Analytic and Translational Genetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

Table 1 | Summary results for primary hypotheses

Hypothesis category	<i>P</i> value (corrected)		Sub-tests of primary hypotheses	Sub-test details	<i>P</i> value (uncorrected)	
a Increased rates of <i>de novo</i> mutations	1.00		Nonsynonymous:synonymous ratio compared to controls ^{7–10,13,14}	Table 2	0.43	
			Loss-of-function:missense ratio compared to controls ^{7–10,13,14}	Table 2	0.37	
	Nonsynonymous	Loss-of-function			Nonsynonymous	Loss-of-function
b Genic recurrence in schizophrenia	0.0007	0.25	Genic recurrence of <i>de novo</i> mutations (current study)	Extended Data Table 2	0.03	0.20
			Enrichment in schizophrenia (literature ^{12–14}) nonsynonymous <i>de novo</i> genes	Table 4, Extended Data Table 5	0.59	0.21
			Increased case/control ¹⁵ ratio of rare (MAF < 0.1%) loss-of-function variants in <i>de novo</i> genes	Ref. 15	0.0003	0.0075
			Excess transmission of nonsynonymous singletons (current study) in <i>de novo</i> genes	—	0.01	0.29
			Enrichment in schizophrenia CNV (literature ^{1,20}) genes	—	0.29	0.66
c Enrichment in candidate genes	0.0098	1.00	Enrichment in ARC/NMDAR genes ²⁰	Table 3, Extended Data Fig. 4	0.0008	0.006
			Enrichment in postsynaptic density genes, excluding ARC/NMDAR genes ²⁰	—	0.24	0.53
			Enrichment in FMRP target genes ⁹	Extended Data Table 3	0.009	0.37
d Enrichment in autism/intellectual disability <i>de novo</i> genes	0.17	0.0055	Enrichment in autism loss-of-function <i>de novo</i> genes ^{6–9}	Table 4, Extended Data Table 5	0.02	0.0007
			Enrichment in intellectual disability loss-of-function <i>de novo</i> genes ^{10,11}	Table 4, Extended Data Table 5	0.27	0.02

Hypotheses are grouped into four broad categories (a–d). Each is comprised of sub-tests from which we derive global evidence for the broad category (see Supplementary Information). **a**, The broad *P* value was generated using Fisher's exact method on the missense, silent and loss-of-function mutation counts. **b**, *P* values were generated using Fisher's combined probability test to combine the sub-tests. **c**, **d**, All genes from each sub-test were combined into a single gene set and enrichment evaluated using the dnenrich software (see main text and Supplementary Information). For **b–d**, we separately evaluated two classes of mutation, nonsynonymous and loss-of-function, making seven tests in total. Corrected *P* values for the broad categories are adjusted by Bonferroni correction for seven tests. *P* values < 0.05 (corrected for broad category tests, uncorrected for sub-tests) in bold.

between the two, we could not confidently distinguish independent parental age effects (Supplementary Information). As expected¹⁶, most *de novo* mutations (79%) we could phase occurred on paternal chromosomes (Supplementary Information). The number of *de novo* mutations per individual followed a Poisson distribution (Extended Data Fig. 3a), in line with previous studies of autism⁶ and schizophrenia¹³. Nevertheless, loss-of-function *de novo* mutations were more common in patients with relatively poor school performance ($P = 0.018$; Extended Data Fig. 3b), but none of the other variables tested—family history, age at onset, gender or having a *de novo* CNV—were significantly associated with mutation rates.

Compared with 731 controls from published data sets (Supplementary Table 2), probands did not have a significant elevation in the relative rates of nonsynonymous to silent mutations, or loss-of-function to missense mutations (Tables 1a and 2). No differences were observed

between schizophrenia cases with or without *de novo* CNVs or between those stratified by common allele risk scores (Extended Data Table 1a). Consistent with their higher loss-of-function mutation rate, those with school grades below the median had significantly elevated loss-of-function to missense ratios compared to both controls ($P = 0.02$) and cases with higher school grades ($P = 0.0095$) (Extended Data Table 1b and Extended Data Fig. 3b). In the absence of an effect of age at onset (that might affect school performance), this suggests loss-of-function mutations occur preferentially in (the large proportion of) schizophrenia cases that have premorbid cognitive impairment¹⁷. All probands attended and graduated from mainstream schools, which excluded people with significant degrees of intellectual disability; moreover, recruiting psychiatrists were explicitly instructed to exclude people with known intellectual disability. Thus, the enrichment of loss-of-function mutations in those with the poorest scholastic attainment cannot be attributed

Table 2 | Ratios of functional classes of *de novo* mutations across various samples

	Controls ^{7–10,13,14}	Current study	Schizophrenia (ref. 14)	Schizophrenia (ref. 13)	Schizophrenia all (refs 12–14)	Autism spectrum disorder ^{6–9}	Intellectual disability ^{10,11}
Nonsynonymous	434	482	68	137	702	789	141
Synonymous	155	155	29	27	211	255	25
Ratio	2.8	3.1	2.3	5.1	3.3	3.1	5.6
<i>P</i> vs controls	—	0.43	0.46	0.0097	0.18	0.41	0.0027
Loss-of-function	49	64	12	20	100	134	34
Missense	376	408	56	113	588	638	104
Ratio	0.13	0.16	0.21	0.18	0.17	0.21	0.33
<i>P</i> vs controls	—	0.37	0.17	0.29	0.17	0.0072	0.0003

Counts of *de novo* mutation in the present study, in previous studies of schizophrenia (refs 13 and 14), and in all studies of schizophrenia combined (Schizophrenia all), which includes this study and an additional small study¹². Controls are unaffected individuals or unaffected siblings of probands with autism spectrum disorder or schizophrenia. To control for factors that influence estimates of absolute rates (sequencing depth, calling, parental age, etc.), we tested for differences between the ratios of classes of *de novo* mutations (nonsynonymous to silent, loss-of-function to missense) in the disorder groups and the controls, using Fisher's exact test. Nominally significant *P* values (< 0.05) are bold.

to the inclusion of individuals with severe intellectual disability, although this does not preclude the presence of individuals with mild intellectual disability among cases with low educational achievement.

Mutations in schizophrenia gene sets

Gene sets selected for independent evidence for relevance to schizophrenia showed enrichment ($P_{\text{corrected}} = 0.0007$) of nonsynonymous *de novo* mutations (Table 1b), indicating that a proportion of mutations are pathogenic for schizophrenia. Specifically, genes were recurrent for *de novo* mutations more than expected (Table 1b, Extended Data Table 2). Genes affected by nonsynonymous *de novo* mutations were also enriched for inherited rare risk alleles (Table 1b), with excess transmission of rare nonsynonymous alleles from parents to the affected probands, as well as enrichment in cases of rare (minor allele frequency (MAF) < 0.001) gene-disruptive mutations in an independent case-control exome sequencing study¹⁵. One gene, *TAF13*, encoding a subunit of the TFIID transcription initiation complex, contains two *de novo* loss-of-function mutations. This recurrence is significant even after genome-wide correction ($P = 1 \times 10^{-6}$; $P_{\text{corrected}} = 0.024$) (Extended Data Table 2). Replication is necessary to firmly establish this as a susceptibility gene.

Mutations enriched in synaptic genes

Previous studies have suggested that CNVs in people with schizophrenia preferentially affect broadly defined sets of synaptic genes^{18,19}. Moreover, a detailed analysis of *de novo* CNVs based on gene sets constructed from experimental proteomics led us to propose that this synaptic enrichment could be explained by mutations affecting proteins closely associated with the *N*-methyl-D-aspartate (NMDA) receptor, which we refer to as the NMDAR complex, and proteins that interact with ARC (activity-regulated cytoskeleton-associated protein), referred to as the ARC complex²⁰. Our primary functional hypothesis in the present study was that genes encoding proteins in the ARC and NMDAR complexes would be disproportionately affected by *de novo* SNV and indel mutations. We additionally postulated that brain-expressed genes that are repressed by fragile X mental retardation protein (FMRP)²¹ would also be enriched for *de novo* mutations because these have been shown to be enriched for *de novo* mutations in ASD⁹. Moreover, FMRP targets include multiple members of the NMDAR and ARC complexes.

We observed experiment-wide significant enrichment for nonsynonymous mutations among the synaptic gene sets (Table 1c), as well as specifically for NMDAR and ARC complexes (Tables 1c and 3, Extended Data Fig. 4). NMDAR and ARC complexes are closely associated elements central to regulating synaptic strength at glutamatergic synapses and have been implicated in cognition. NMDA signalling triggers multiple processes required for inducing synaptic plasticity²². ARC is involved in almost all known forms of synaptic plasticity, including

synaptic remodelling, the consolidation of changes in synaptic strength linked to memory and response to stress^{23–25}, and regulating synapse elimination during development²⁶, a process believed to be aberrant in schizophrenia²⁷.

FMRP targets were also enriched for nonsynonymous *de novo* mutations (Table 1c), even after NMDAR, ARC and the broader group of postsynaptic density genes were removed (Extended Data Table 3). Given that loss of FMRP results in widespread deficits in synaptic plasticity²⁸, these findings again implicate pathogenic disruption of plasticity mechanisms in schizophrenia. Secondary analyses to dissect the FMRP target enrichment by subdividing genes by gene ontology²⁹ membership did not identify significant categories.

Support for the candidate hypotheses were replicable and robust to study design. In the schizophrenia case-control study¹⁵, rare (MAF < 0.001) loss-of-function mutations were enriched in NMDAR ($P = 0.02$), ARC ($P = 1 \times 10^{-3}$), and FMRP target ($P = 0.003$) sets. Across studies, loss-of-function enrichments in the ARC complex were particularly striking; 17-fold here (Table 3 and Extended Data Fig. 4f) and 19-fold in the case-control study, indicating that disruption of ARC function has particularly strong effects on disease risk.

Aiming to identify hitherto unsuspected disease mechanisms, we undertook an hypothesis-free analysis based on the comprehensive gene ontology (GO) annotations²⁹. A single category (GO:0051017) was significantly enriched for nonsynonymous *de novo* mutations ($P = 6.6 \times 10^{-6}$) after correction for all GO categories ($P_{\text{corrected}} = 0.032$). Genes in GO:0051017, assembly of actin filament bundles, are intimately involved in synaptic plasticity, and are functionally interconnected with ARC and NMDAR signalling (see Supplementary Information). Even after removal of genes overlapping with ARC/NMDAR sets, GO:0051017 remained enriched eightfold for mutations ($P = 0.0011$). Although not significant in the case-control data set¹⁵, this category was significantly enriched for *de novo* CNVs in a study of ASD³⁰. It also includes *KCTD13*, the gene responsible for some of the phenotypes associated with CNVs at 16p11.2 (ref. 31), duplication of which is a risk factor for schizophrenia⁴. *KCTD13* also maps to a schizophrenia genome-wide significant SNP locus³².

Connectivity of mutated synaptic genes

Seeking further insights into synaptic pathology, we identified interactions involving proteins with *de novo* mutations using a synaptic interactome database³³ (Supplementary Information). Proteins with nonsynonymous *de novo* mutations had more connectivity than expected among each other (Fig. 1a) and with synaptic proteins in general, indicating a greater than average importance to the synapse. Directly interacting proteins with *de novo* mutations are involved in multiple processes regulating synaptic plasticity, particularly NMDA, AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) and kainate

Table 3 | Enrichment of *de novo* mutations in postsynaptic protein complexes

		Current study								Schizophrenia (ref. 14)		Schizophrenia (ref. 13)		Schizophrenia all (refs 12–14)		Autism spectrum disorder ^{6–9}		Intellectual disability ^{10,11}	
		Nonsynonymous (482)				Loss-of-function (64)				Non-synony-mous (68)	Loss-of-function (12)	Non-synony-mous (137)	Loss-of-function (20)	Non-synony-mous (702)	Loss-of-function (100)	Non-synony-mous (789)	Loss-of-function (134)	Non-synony-mous (141)	Loss-of-function (34)
Gene set	Genes (N)	<i>P</i>	No. mut.	O/E	<i>P</i>	No. mut.	O/E	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	
Postsynaptic density	681	0.019	34	1.46	0.091	6	1.92	0.84	0.45	0.65	0.64	0.091	0.12	0.47	0.064	0.0015	0.00004		
ARC complex	28	0.00048	6	6.06	0.005	2	17.42	1	1	1	1	0.0035	0.015	0.22	0.22	0.00002	0.0015		
NMDAR complex	60	0.025	6	2.74	0.035	2	6.99	1	1	0.13	0.086	0.016	0.011	0.031	0.46	0.00002	0.00002		

Statistical significance for enrichment of *de novo* mutations in glutamatergic postsynaptic gene sets²⁰. Nominally significant *P* values (< 0.05), as calculated with dnenrich (see Supplementary Information), are marked in bold. No. mut., mutation counts in each set. O/E, observed-to-expected ratio of mutational hits (fold-enrichment statistic) calculated with dnenrich. Samples and classes of mutations are as Table 2. Total numbers of mutations for each class in each sample are given in parentheses. Additional details for the current study, including genes and 95% credible intervals (CI) for the O/E statistics, are given in Extended Data Fig. 4.

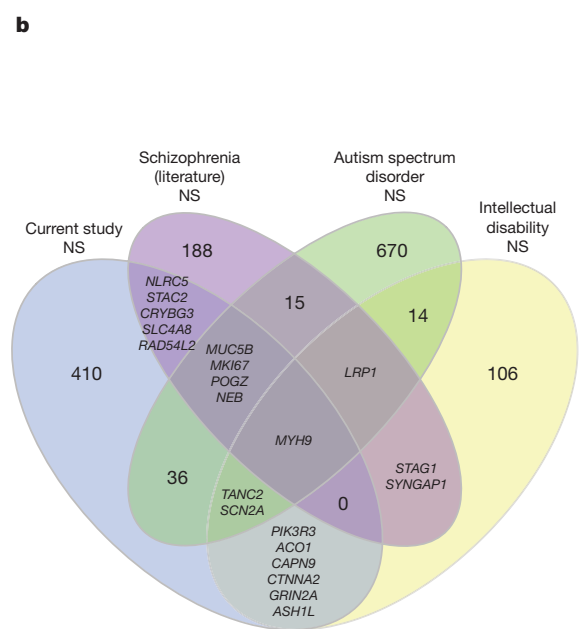
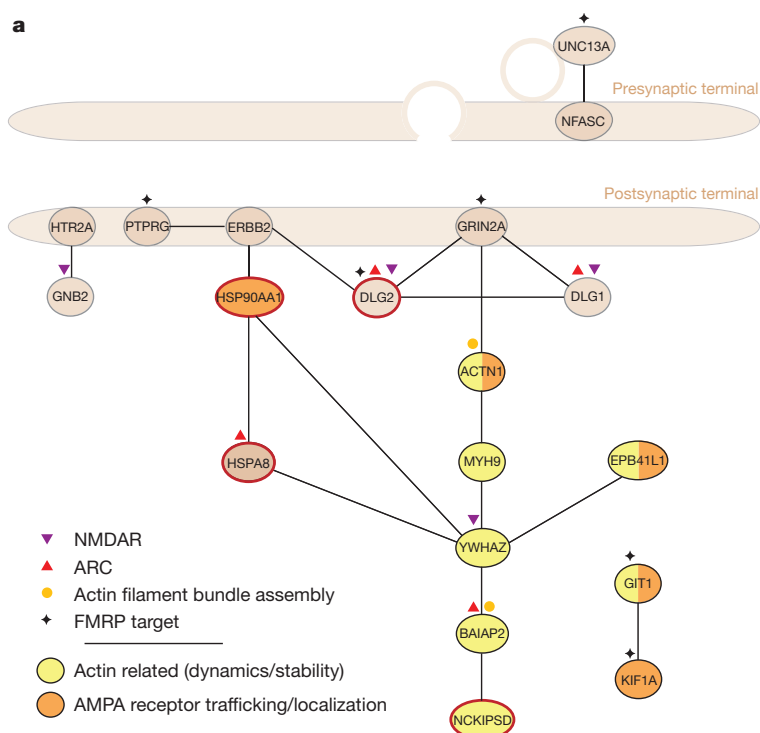


Figure 1 | *De novo* mutations from schizophrenia affect genes coding for synaptic proteins and genes affected in other neuropsychiatric diseases.

a, Synaptic protein–protein interactions between proteins affected by nonsynonymous *de novo* mutations in schizophrenia. Interactions were retrieved from the expert-curated lists in the SynSysNet database (<http://bioinformatics.charite.de/synsysnet/>) and plotted to show their general pre/postsynaptic localization. Proteins belonging to various functional sets are as marked, and the four proteins encoded by genes with loss-of-function

receptor trafficking, and the regulation of actin dynamics. These interactions involve genes not present in our pre-assigned NMDAR/ARC and actin filament complexes (Supplementary Information). Although our analyses highlighted postsynaptic processes, some of the interacting synaptic proteins with *de novo* mutations are presynaptic (Fig. 1a, Supplementary Information and Extended Data Fig. 4a). Pre- and postsynaptic proteins are, however, closely functionally related; indeed, trans-synaptic effects of presynaptic proteins on the regulation of AMPA receptor trafficking and NMDAR-dependent plasticity have recently been described³⁴.

We were unable to replicate a previous report of prenatal bias in brain expression for genes with schizophrenia *de novo* mutations¹³ using microarray or RNA-seq data (Supplementary Information and Extended Data Table 4).

mutations are noted with a red outline. Proteins with nonsynonymous *de novo* mutations had more than expected direct interconnections ($P = 0.008$), which was consistent with more overall connectivity to synaptic proteins as a whole ($P = 0.005$). **b**, Overlap of genes bearing nonsynonymous (NS) *de novo* mutations in schizophrenia (refs 12–14), autism spectrum disorder (refs 6–9) and intellectual disability (refs 10, 11). Overlaps of six or fewer genes are listed by name. See Extended Data Table 5 for statistical significance of these overlaps, and see Table 2 and text for disease sets.

Overlaps between disorders

CNV loci associated with schizophrenia overlap with those seen in ASD, intellectual disability and attention deficit hyperactivity disorder^{1,4,35}. However, because pathogenic CNVs typically span multiple genes and are concentrated in a relatively small fraction of the genome³⁶, it is possible that this may not indicate cross-disorder effects at the level of specific genes. Therefore, we sought evidence for shared genetic aetiology between schizophrenia and both intellectual disability and ASD³⁷ by testing for overlap of genes affected by *de novo* mutations in schizophrenia, ASD and intellectual disability.

Genes with *de novo* mutations in the current study overlapped those affected by *de novo* mutations in ASD^{6–9} and intellectual disability^{10,11} (Fig. 1b and Tables 1d and 4), but not in controls (Extended Data Table 5). Moreover, loss-of-function mutations in schizophrenia were enriched

Table 4 | Overlap between genes hit by *de novo* mutations in this study and other phenotypes

Gene set	Mutation class (N genes)	P	Current study (mutations)		P	No. mut.
			Nonsynonymous (482)	Loss-of-function (64)		
Schizophrenia (ref. 14)	Nonsynonymous (67)	0.22	6	0.021	3	
	Loss-of-function (12)	0.051	3	0.11	1	
Schizophrenia (ref. 13)	Nonsynonymous (136)	0.79	5	1	0	
	Loss-of-function (20)	0.24	2	1	0	
Autism spectrum disorder ^{6–9}	Nonsynonymous (743)	0.14	45	0.023	9	
	Loss-of-function (128)	0.015	11	0.00072	4	
Intellectual disability ^{10,11}	Nonsynonymous (132)	0.032	9	0.031	1	
	Loss-of-function (30)	0.27	1	0.019	1	
Controls ^{7–10,13,14}	Nonsynonymous (424)	0.59	21	1	0	
	Loss-of-function (49)	0.6	2	1	0	

Number of mutations (No. mut.) in present study in gene sets derived from previous studies. *P* values are calculated with dnenrich for enrichment of mutations in the gene sets from previous studies (see Supplementary Information). Nominally significant *P* values (<0.05) are in bold. Disease sets and mutation classes are as Table 2. Additional comparisons are given in Extended Data Table 5.

even in the small subset of genes ($N = 7$) with recurrent loss-of-function *de novo* mutations in ASD ($P = 0.0018$) or intellectual disability ($P = 0.019$), the mutations occurring in *SCN2A* (encoding an alpha subunit of voltage-gated sodium channels, a major mediator of neuronal firing and action potential propagation) and *POGZ* (whose involvement in mitosis suggests a possible role in regulating neuronal proliferation³⁸). Both *SCN2A* and *POGZ* are now established ASD genes³⁹. Other notable genes affected by loss-of-function mutations in the present study for which there is prior support for loss-of-function mutations in other neurodevelopmental disorders include *DLG2* and *SHANK1* (Supplementary Information). Thus, we now show overlap between schizophrenia, ASD and intellectual disability at the resolution not just of loci or even individual genes, but even of mutations with similar functional (loss-of-function) effects.

ARC/NMDAR complexes (Table 3) and FMRP targets (Extended Data Table 3) were enriched for *de novo* mutations in intellectual disability, and NMDAR and FMRP targets were also enriched in ASD, providing further evidence of shared disease mechanisms. However, we also find differences between the disorders. In general, enrichment statistics were stronger for ASD and intellectual disability than schizophrenia, particularly for loss-of-function mutations (Table 2), despite the relatively small number of intellectual disability trios. Genes and mutation sites were most highly conserved in intellectual disability, then ASD, with schizophrenia least conserved (Supplementary Information and Extended Data Table 6). These findings indicate that highly disruptive mutations have a relatively lesser role in schizophrenia, and also that the disorders differ by severity of functional impairment, consistent with the hypothesis of an underlying gradient of neurodevelopmental pathology⁴⁰ indexed by cognitive impairment, with intellectual disability at one extreme.

That the most damaging mutations reflect a gradient of neurodevelopmental impairment is further supported by the observation that, within schizophrenia, the highest rate of loss-of-function mutations (Extended Data Fig. 3b) occurred in individuals likely to have the greatest cognitive impairment (lowest scholastic attainment), as does the observation that the loss-of-function genic overlap between schizophrenia and both autism and intellectual disability is dependent on the *de novo* mutations (including *SCN2A* and *POGZ*) in those individuals (Extended Data Table 1c). However, as noted above, the enrichment of loss-of-function mutations in those with the poorest scholastic attainment cannot be attributed to the inclusion of individuals with severe intellectual disability. Moreover, when we exclude cases with low scholastic attainment, we still see significant enrichment of the synaptic pathways that are enriched in the full sample (Supplementary Information and Extended Data Table 1c). Thus, our implication of synaptic protein complexes is not dependent on mutations present in a subset of cases with severely impaired cognitive function.

Discussion

In the largest exome-sequencing-based study of *de novo* mutations in schizophrenia, we demonstrate a convergence of *de novo* mutations on multiply defined sets of functionally related proteins, pointing to the regulation of plasticity at glutamatergic synapses as a pathogenic mechanism in schizophrenia. How disruption of these synaptic mechanisms affects brain function to produce psychopathology cannot be answered by genetic studies alone, but our identification of *de novo* mutations in these gene sets provides the basis to address this. Our findings of overlaps between the pathogenic mechanisms underlying schizophrenia and those in autism and intellectual disability lend support to recent, controversial suggestions that our understanding of these disorders might be advanced better by research that integrates findings across multiple disorders and places more emphasis on domains of psychopathology (for example, cognition) and their neurobiological substrates rather than current diagnostic categories^{40,41}.

METHODS SUMMARY

Parent proband trios ($N = 623$), where the proband had a history of hospitalization for schizophrenia or schizoaffective disorder, were recruited from psychiatric hospitals in Bulgaria. Proband attend mainstream schools which excluded people with intellectual disability; all graduated with a pass. Exome DNA was captured from genomic DNA (whole blood), using either Agilent or NimbleGen array-based capture, and subjected to paired-end sequencing on Illumina HiSeq sequencers. The BWA-Picard-GATK pipeline was used for sequence alignment and variant calling. Putative *de novo* mutations were identified using Plink/Seq (<http://atgu.mgh.harvard.edu/plinkseq>) and were validated using Sanger sequencing. We used Plink/Seq to annotate mutations according to RefSeq gene transcripts (UCSC Genome Browser, <http://genome.ucsc.edu>). Mutation rate was tested for association with clinical and other covariates using a generalized linear model. Rates of functional classes of mutations in probands were compared with those in published controls using Fisher's exact test. Mutations were tested for recurrence, enrichment in candidate gene sets, and enrichment in genes affected by *de novo* mutations in previous studies using the dnenrich software (Supplementary Information, <https://bitbucket.org/statgen/dnenrich>). dnenrich calculates one-sided P values under a binomial model of greater than expected hits using randomly placed mutations accounting for gene size, sequencing coverage, tri-nucleotide contexts and functional effects of the observed mutations. Candidate gene sets and studies of neuropsychiatric disease are described in the main and Supplementary Information. Primary hypotheses (Table 1) were Bonferroni-corrected for multiple testing.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 July; accepted 3 December 2013.

Published online 22 January; corrected online 12 February 2014 (see full-text HTML version for details).

- Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev. Genet.* **13**, 537–551 (2012).
- Bundy, H., Stahl, D. & MacCabe, J. H. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatr. Scand.* **123**, 98–106 (2011).
- McClellan, J. M., Sussner, E. & King, M.-C. Schizophrenia: a common disease caused by multiple rare alleles. *Br. J. Psychiatry* **190**, 194–199 (2007).
- Malhotra, D. & Sebat, J. Genetics: Fish heads and human disease. *Nature* **485**, 318–319 (2012).
- Rees, E., Moskvina, V., Owen, M. J., O'Donovan, M. C. & Kirov, G. De novo rates and selection of schizophrenia-associated copy number variants. *Biol. Psychiatry* **70**, 1109–1114 (2011).
- Neale, B. M. et al. Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
- O'Roak, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
- Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- Iossifov, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- Rauch, A. et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
- de Ligt, J. et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
- Girard, S. L. et al. Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genet.* **43**, 860–863 (2011).
- Xu, B. et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature Genet.* **44**, 1365–1369 (2012).
- Gulsuner, S. et al. Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
- Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* <http://dx.doi.org/10.1038/nature12975> (22 January 2014).
- Kong, A. et al. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
- Meshulam-Gately, R. I., Giuliano, A. J., Goff, K. P. & Faraone, S. V. & Seidman, L. J. Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology* **23**, 315–336 (2009).
- Glessner, J. T. et al. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl Acad. Sci. USA* **107**, 10584–10589 (2010).
- Malhotra, D. et al. High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia. *Neuron* **72**, 951–963 (2011).
- Kirov, G. et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
- Darnell, J. C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Malenka, R. C. & Nicoll, R. A. NMDA-receptor-dependent synaptic plasticity: multiple forms and mechanisms. *Trends Neurosci.* **16**, 521–527 (1993).

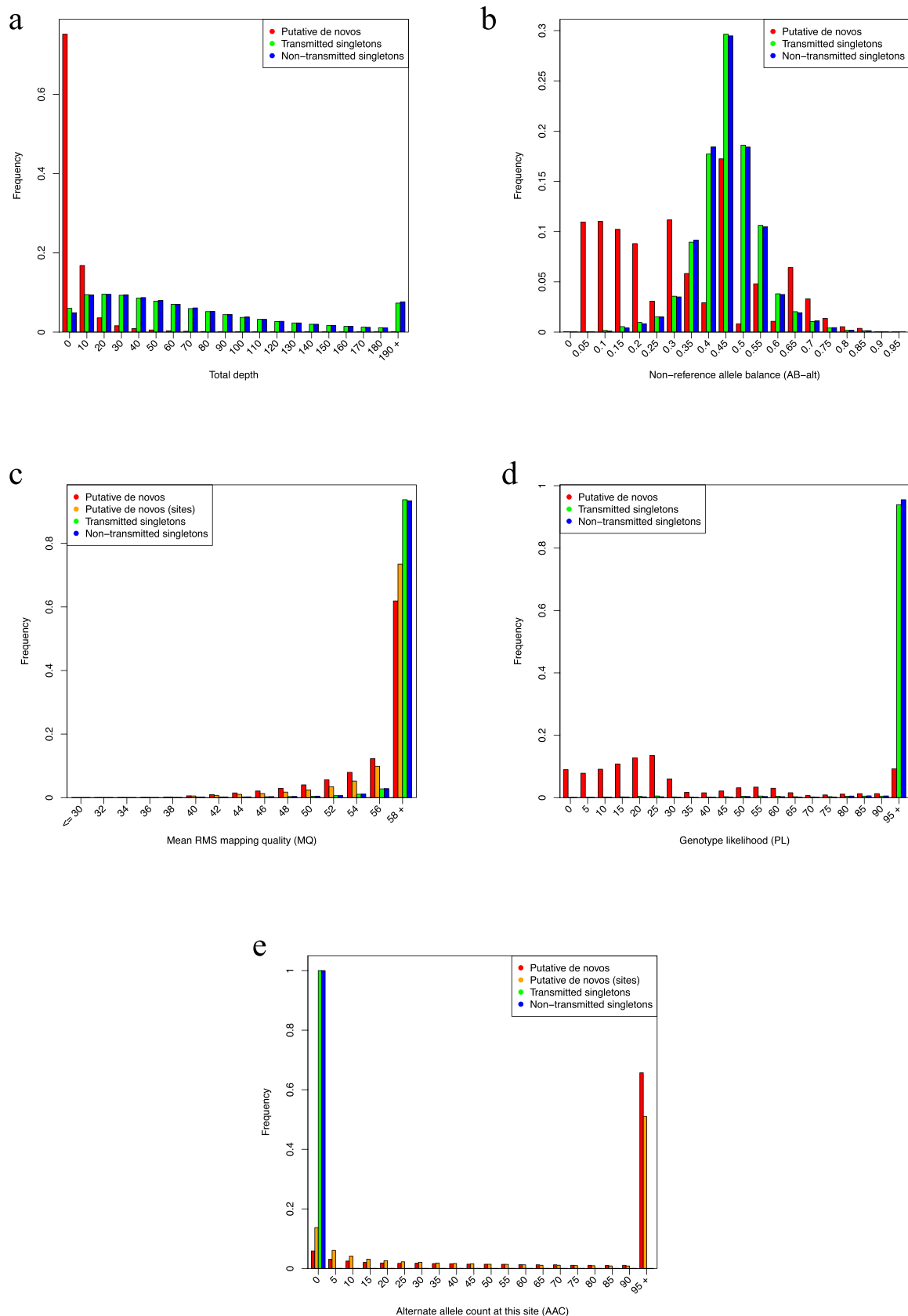
23. Bramham, C. R. *et al.* The Arc of synaptic memory. *Exp. Brain Res.* **200**, 125–140 (2010).
24. Korb, E. & Finkbeiner, S. Arc in synaptic plasticity: from gene to behavior. *Trends Neurosci.* **34**, 591–598 (2011).
25. Shepherd, J. D. & Bear, M. F. New views of Arc, a master regulator of synaptic plasticity. *Nature Neurosci.* **14**, 279–284 (2011).
26. Mikuni, T. *et al.* Arc/Arg3.1 Is a postsynaptic mediator of activity-dependent synapse elimination in the developing cerebellum. *Neuron* **78**, 1024–1035 (2013).
27. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17**, 319–334 (1982).
28. Darnell, J. C. & Klann, E. The translation of translational control by FMRP: therapeutic targets for FXS. *Nature Neurosci.* **16**, 1530–1536 (2013).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
30. Gilman, S. R. *et al.* Rare *de novo* variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* **70**, 898–907 (2011).
31. Golzio, C. *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).
32. Steinberg, S. *et al.* Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum. Mol. Genet.* **20**, 4076–4081 (2011).
33. von Eichborn, J. *et al.* SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations. *Nucleic Acids Res.* **41**, D834–D840 (2013).
34. Aoto, J., Martinelli, D. C., Malenka, R. C., Tabuchi, K. & Südhof, T. C. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. *Cell* **154**, 75–88 (2013).
35. Williams, N. M. *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* **376**, 1401–1408 (2010).
36. Girirajan, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med.* **367**, 1321–1331 (2012).
37. Owen, M. J., O'Donovan, M. C., Thapar, A. & Craddock, N. Neurodevelopmental hypothesis of schizophrenia. *Br. J. Psychiatry* **198**, 173–175 (2011).
38. Nozawa, R.-S. *et al.* Human POGZ modulates dissociation of HP1 α from mitotic chromosome arms through Aurora B activation. *Nature Cell Biol.* **12**, 719–727 (2010).
39. Buxbaum, J. D. *et al.* The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* **76**, 1052–1056 (2012).
40. Craddock, N. & Owen, M. J. The Kraepelinian dichotomy – going, going...but still not gone. *Br. J. Psychiatry* **196**, 92–95 (2010).
41. Insel, T. R. Transforming diagnosis. *NIMH Director's Blog* <http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml> (29 April 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Work in Cardiff was supported by Medical Research Council (MRC) Centre (G0800509) and Program Grants (G0801418), the European Community's Seventh Framework Programme (HEALTH-F2-2010-241909 (Project EU-GEI)), and NIMH (2 P50 MH066392-05A1). Work at the Icahn School of Medicine at Mount Sinai was supported by the Friedman Brain Institute, the Institute for Genomics and Multiscale Biology (including computational resources and staff expertise provided by the Department of Scientific Computing), and National Institutes of Health grants R01HG005827 (S.M.P.), R01MH099126 (S.M.P.), and R01MH071681 (P.S.). Work at the Broad Institute was funded by Fidelity Foundations, the Sylvan Herman Foundation, philanthropic gifts from K. and E. Dauten, and the Stanley Medical Research Institute. Work at the Wellcome Trust Sanger Institute was supported by The Wellcome Trust (grant numbers WT089062 and WT098051) and also by the European Commission FP7 project gEUVADIS no. 261123 (P.P.). We would like to thank M. Daly, B. Neale and K. Samocha for discussions and providing unpublished autism data. We would also like to acknowledge M. DePristo, S. Gabriel, T. J. Fennel, K. Shakir, C. Tolonen and H. Shah for their help in generating and processing the various data sets.

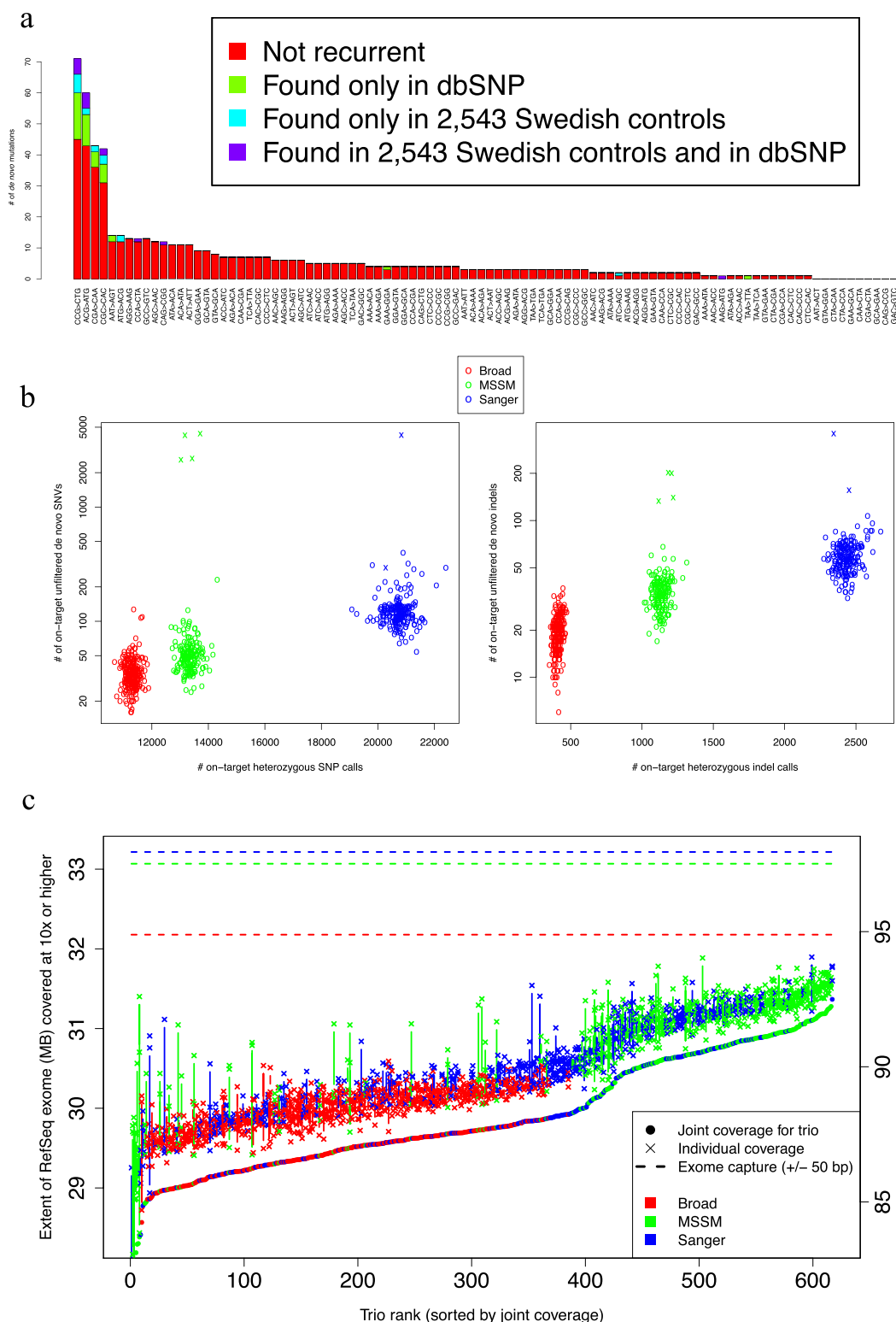
Author Contributions The project was led in Cardiff by M.C.O.D. and M.J.O., in Mount Sinai by S.M.P. and P.S., at the Broad by S.A.M. and J.L.M., and at the Sanger by A.P.; H.J.W., J.L.M., K.C., J.S.J., D.D.B., M.M. and S.A.R. were responsible for sample processing and data management. M.F., H.J.W., P.G., D.M.R., D.H.K., G.K., E.R. and S.D. processed NGS data, annotated and validated mutations. L.G., N.C., I.H., S.D., H.J.W. and S.A.R. undertook validation of mutations and additional lab work. A.J.P., M.F., D.H.K., S.M.P. and P.H. co-ordinated/undertook the main bioinformatics/statistical analyses. E.R., D.M.R., E.B., P.P., E.H. and P.R. performed additional analyses. S.G.G. contributed additional insights into synaptic biology. Sample recruitment was led by G.K. and V.M.; The main findings were interpreted by M.C.O.D., M.F., M.J.O., P.H., G.K., E.M.S., S.A.M., D.H.K., A.J.P., A.P., S.M.P. and P.S. who drafted the manuscript.

Author Information Data included in this manuscript have been deposited at dbGaP under accession number phs000687.v1.p1 and is available for download at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000687.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.O. (owenmj@cardiff.ac.uk).



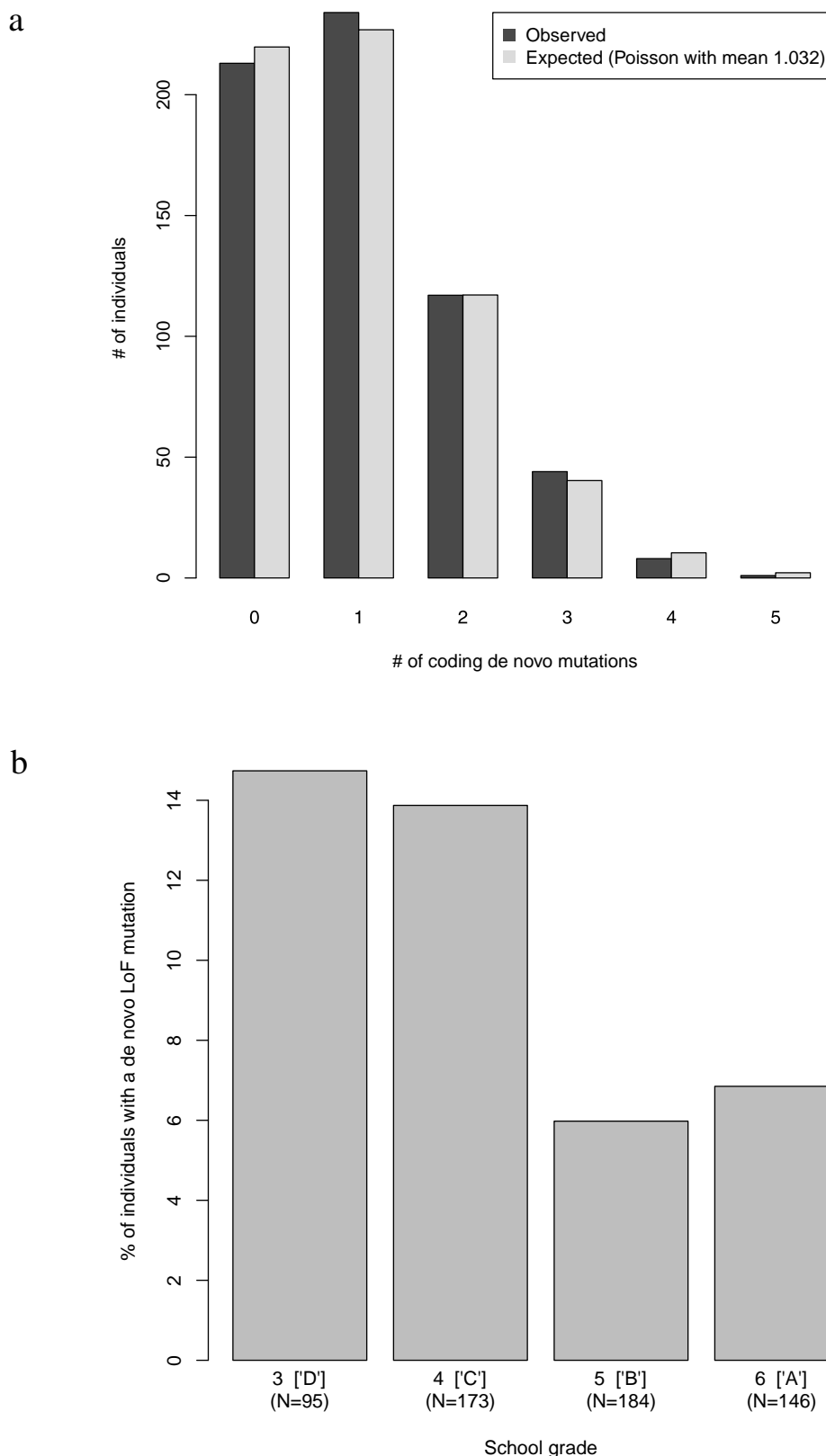
Extended Data Figure 1 | Comparison of sequencing metrics for putative *de novo* calls and parental singletons. **a–e**, Putative *de novo* calls (child heterozygous, both parents homozygous reference; N parent-proband trios = 623) were compared with variants observed in only a single parent (“singletons”), in terms of depth of all reads at the variant site (**a**), fraction of reads with the alternate allele (AB = allele balance) (**b**), mapping quality of the

reads at the site (MQ) (**c**), the likelihood of the heterozygous genotype (PL = Phred-scaled likelihood) (**d**), and the number of other samples in the present study with a non-reference allele at that site (AAC = alternate allele count) (**e**). Distributions were calculated for putative *de novo* variants (red), or grouped by sites of putatively recurrent *de novo* mutations (orange) when relevant, transmitted singletons (green), and non-transmitted singletons (blue).



Extended Data Figure 2 | Metrics for *de novo* variants across cohorts and trios. **a**, Rates of recurrence of validated *de novo* mutations for tri-nucleotide sequences. For each of 96 possible tri-nucleotide base contexts of single-base mutations (accounting for strand symmetry by reverse complementarity), the number of observed *de novo* SNV is plotted (sorted by this count). Mutation counts are sub-divided into those not found in external data (red), those found in dbSNP (build 137, green), those found in controls ($N = 2543$) in the parallel exome sequencing study¹⁵ (cyan), and those found both in dbSNP and that study (purple). **b**, Comparison of on-target heterozygous SNV and indel call rate with putative *de novo* mutation calls. For each proband ($N = 623$), the number of heterozygous SNV and indel calls is compared with the number of putative *de novo* mutations (child heterozygous, both parents homozygous

reference). Proband is coloured by sequencing centre (see Supplementary Information for differences in exome capture), and six trios are noticeable outliers from all others (marked by 'x') in terms of number of putative *de novo* mutations. **c**, Variation in sequencing coverage between and across trios and sequencing centres. For each trio ($N = 623$), the number of bases covered by 10 reads or more for each member (marked by 'x') and the joint coverage⁹ in all three members (marked by points) are plotted at corresponding horizontal points; trios are sorted in increasing order of joint coverage and coloured by sequencing centre (see Supplementary Information). The intersection of each exome capture with the RefSeq coding sequence is marked by respective dotted lines.



Extended Data Figure 3 | *De novo* mutation counts and rates.

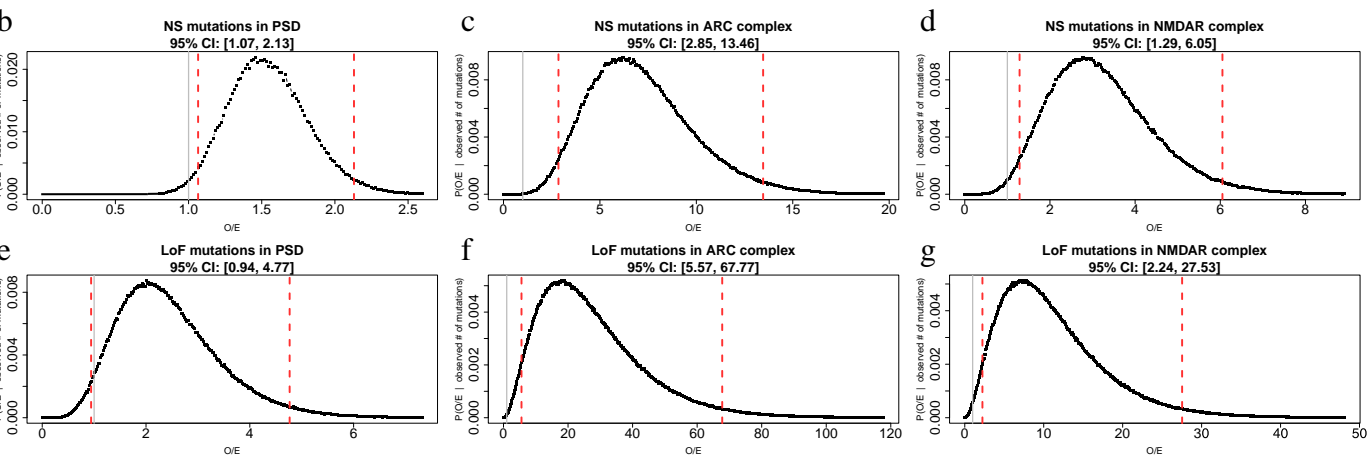
a, The observed distribution of number of validated RefSeq-coding (see Supplementary Information) *de novo* mutations found for each trio ($N = 617$) is compared with that expected from a Poisson distribution with a rate equal to the observed mean number of *de novo* mutations ($\lambda = 1.032$). **b**, Deleterious mutation rate inversely correlates with academic performance. Individuals were grouped according to their final school grade (3–6, corresponding to D, C,

B, A in the US system, <http://www.fulbright.bg/en/p-Educational-System-of-Bulgaria-18/>), and the proportion of individuals with one or more *de novo* loss-of-function mutations is plotted. N , number of individuals in each group. See Supplementary Information for details on linear regression performed to evaluate association; note that 19 samples were removed from this analysis for missing parental age or school grade information, leaving a total of 598 trios.

a

Gene set	# genes	Mut type	# mut hitting set	p-value	Genes hit (counts)	<i>de novo</i> *CNV p-value (Kirov et al., 2012)	<i>de novo</i> CNV genes	Case-control CNVs p-value (Kirov et al., 2012)
PSD	681	NS	34	0.019	ACTN1, ANK1, BAIAP2(x2), BRSK1, CAPN5, DLG1, DLG2, EPB41, EPB41L1, FARS1, GIT1, GNB2, HSP90AA1, HSPA8(x2), ITSN1, KIF1A, MYH11, MYH9, MYO18A, NCKIPSD, NFASC, NRXN1, PLXNA1, PTK2B, RIMS1, SHANK1, SLC25A12, SNTD1, SORBS2, SRCIN1, UNC13A, YWHAZ	4.50E-02	ALDOA, CYFIP1, DLG1, DLG2, DLGAP1, HSPB1, MAPK3, MDH2, RPH3A, RYR2, SNTD1, STX1A, TAOK2, TJP1, YWHAZ	-
		LoF	6	0.091	DLG2, HSP90AA1, HSPA8, ITSN1, NCKIPSD, SHANK1			
ARC complex	28	NS	6	0.0005	BAIAP2(x2), DLG1, DLG2, HSPA8(x2)	2.51E-04	CYFIP1, DLG1, DLG2, DLGAP1	0.14
		LoF	2	0.005	DLG2, HSPA8			
NMDAR complex	60	NS	6	0.025	DLG1, DLG2, GNB2, PTK2B, SHANK1, YWHAZ	6.30E-03	DLG1, DLG2, DLGAP1, MAPK3, STX1A, TJP1, YWHAZ	1.50E-03
		LoF	2	0.035	DLG2, SHANK1			

* compared to control *de novo* CNVs



Extended Data Figure 4 | Enrichment of *de novo* SNVs, indels and CNVs in genes encoding postsynaptic complexes at glutamatergic synapses.

a, Number of *de novo* mutations (*N* cases = 617) in postsynaptic complexes in current study (and genes affected) are shown alongside the most conservative estimate of *de novo* CNV enrichment from ref. 20 (*N* cases studied = 662). NS, nonsynonymous, LoF, loss-of-function, PSD, postsynaptic density. The NMDAR complex gene set was derived a priori from a published proteomics data set⁴². To avoid investigator bias, we did not add additional members *post hoc*, thus omitting genes with *de novo* mutations and important NMDAR functions; these include *GRIN2A*, which encodes a subunit of the NMDA receptor itself, and *AKAP9*, which directly anchors protein complexes involved in signalling at NMDA receptors⁴³. *P* < 0.05 are marked in bold as are

genes hit by mutations in the current study and by *de novo* CNVs in ref. 20. b–g, 95% credible intervals (CI) for fold-enrichment statistics of *de novo* mutations in postsynaptic gene sets (corresponding to enrichments in a, above, and as marked) were calculated from the posterior distributions of fold-enrichment (O/E, observed to expected) statistic values for individuals in this study. Point estimates of O/E are given in Table 3, and correspond to the distribution modes here. The 95% CI is marked by red vertical lines, and a null effect size (value of 1) is marked by a grey line. Note that loss-of-function mutations in the large postsynaptic density set are not significantly enriched, and thus the corresponding CI includes an effect size of 1. All posterior distributions were calculated using dnenrich, as described in the Supplementary Information.

42. Pocklington, A. J., Armstrong, J. D. & Grant, S. G. N. Organization of brain complexity–synapse proteome form and function. *Brief. Funct. Genomics Proteomics* **5**, 66–73 (2006).

43. Coba, M. P. et al. TNiK is required for postsynaptic and nuclear signaling pathways and cognitive function. *J. Neurosci.* **32**, 13987–13999 (2012).

Extended Data Table 1 | Stratification of *de novo* mutations based on polygenic burden, presence of a 'pathogenic' CNV, or poor scholastic achievement**a**

	Probands with top 50% of polygenic scores	Probands with bottom 50% of polygenic scores	Probands with top 50% of polygenic scores or a 'pathogenic' CNV	Probands with bottom 50% of polygenic scores and no 'pathogenic' CNV
NS	210	229	228	214
S	71	66	74	63
Ratio	2.96	3.47	3.08	3.4
<i>P</i>	0.43		0.63	

LoF	24	29	26	27
missense	182	196	198	183
Ratio	0.13	0.15	0.13	0.15
<i>P</i>	0.77		0.77	

b

	Controls ^{7-10,13-14}	Probands with poor scholastic performance (school grades 3 or 4)	Probands with high scholastic performance (school grades 5 or 6)
NS	434	222	242
S	155	67	84
Ratio	2.8	3.3	2.9
<i>P</i> vs poor scholastic performance	0.32	-	0.51

LoF	49	40	23
missense	376	177	214
Ratio	0.13	0.23	0.11
<i>P</i> vs poor scholastic performance	0.021	-	0.0095

c

Gene set	Genes (N)	All probands				Exclude probands with lowest school grade (3)				Exclude probands with a 'pathogenic' CNV or with a polygenic score in the top 5% among probands			
		NS (482)		LoF (64)		NS (398)		LoF (49)		NS (423)		LoF (54)	
		<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut
PSD	681	0.019	34	0.091	6	0.0033	32	0.031	6	0.039	29	0.047	6
ARC complex	28	0.00048	6	0.005	2	0.0002	6	0.0036	2	0.0019	5	0.0048	2
NMDAR complex	60	0.025	6	0.035	2	0.036	5	0.02	2	0.045	5	0.026	2
FMRP targets	784	0.0094	64	0.37	7	0.0041	56	0.28	6	0.031	54	0.55	5
actin filament bundle assembly	34	6.57E-06	8	1	0	0.0023	5	1	0	0.0005	6	1	0
autism LoF genes	128	0.015	11	0.00072	4	0.41	6	0.52	1	0.025	9	0.0013	3
ID LoF genes	30	0.27	1	0.019	1	1	0	1	0	0.22	1	0.016	1

a, Ratios of nonsynonymous to synonymous (NS:S) and loss-of-function to missense (LoF:missense) *de novo* mutations were compared (Fisher's exact test) between those found in individuals with a high polygenic score (top 50%) and those in the individuals in the bottom 50% of the polygenic score distribution (scores previously generated for this sample⁴⁴). Individuals were additionally split based on the presence of a 'pathogenic' CNV. A 'pathogenic' CNV was defined as either a *de novo* CNV identified for these samples in ref. 20, or a CNV associated with psychiatric disease¹. *P* values were computed using Fisher's exact test as in Table 2. **b**, Ratios of NS:S and LoF:missense *de novo* mutations were compared (Fisher's exact test) between schizophrenia probands with poor scholastic performance and (1) controls; (2) probands with high scholastic performance. Nominally significant results ($P < 0.05$) are marked in bold. **c**, Enrichment of *de novo* mutations (as calculated by dnenrich, see Supplementary Information) including all individuals, after excluding individuals with the lowest scholastic achievement (a school grade of 3), or excluding those with a 'pathogenic' CNV (see above) or a polygenic score in the top 5% of the distribution (see above). These secondary exclusion analyses were performed on those gene sets identified as significant in the analyses of the full set; statistics shown as in Table 3. Total numbers of nonsynonymous (NS) and loss-of-function (LoF) mutations in each subset of probands are given in parentheses, and $P < 0.05$ are marked in bold. PSD, postsynaptic density.

44. Ruderfer, D. M, et al. A family-based study of common polygenic variation and risk of schizophrenia. *Mol. Psychiatry* **16**, 887–888 (2011).

Extended Data Table 2 | Genes overlapped by two nonsynonymous *de novo* mutations in schizophrenia probands

18 genes with recurrent nonsynonymous <i>de novo</i> s mutations ($p = 0.0314$)				
Gene	De novo mutations	Nominal p -value for recurrence of NS (LoF) <i>de novo</i> s	Case/control counts of rare (MAF < 0.001) LoF mutations in Purcell, et al. ¹⁵	Nominal case/control p -value
AKD1	frameshift, missense	0.0024	2/8	1
BAIAP2	codon-deletion, missense	0.00042	1/0	0.53
C7orf60	missense (x2)	0.00013	0/0	1
CD14	missense (x2)	0.00021	0/0	1
HSPA8	frameshift, missense	0.00035	0/0	1
HUWE1	missense (x2)	0.014	0/0	1
KIAA1244	missense (x2)	0.0041	0/0	1
KIF18A	missense (x2)	0.00063	1/0	0.52
LPHN2	missense, nonsense	0.0014	0/0	1
MUC6	missense (x2)	0.0059	3/5	1
NIPAL3	missense, nonsense	0.00017	0/0	1
NLRC5	missense (x2)	0.0025	3/4	1
PHC2	missense (x2)	0.00072	0/0	1
PHF7	missense, nonsense	9.80E-05	0/0	1
PIK3C2B	frameshift, missense	0.0024	3/0	0.11
PSPC1	missense, nonsense	0.00034	0/0	1
RYR3	missense (x2)	0.018	4/1	0.22
TAF13	frameshift, nonsense	1.5e-05 (1.2e-06)	1/0	0.53

Genes hit by nonsynonymous (NS) mutations in two different probands with schizophrenia ($N = 18$) are listed, with the expected functional impact of those mutations and the nominal P value for genic recurrence (calculated by dnenrich); for the single instance of two loss-of-function alleles in a single gene (*TAF13*), the P value for loss-of-function recurrence is given in parentheses; this is bolded since it is significant after Bonferroni correction for multiple testing of all genes (see Supplementary Information). Also shown are the case/control counts from the parallel exome sequencing study¹⁵ and the corresponding nominal P value for association with schizophrenia.

Extended Data Table 3 | Enrichment of *de novo* mutations in genes targeted by FMRP and conditional analysis of enrichment in postsynaptic density complexes

Genes tested	Mutations Genes (N)	Current study				Schizophrenia (ref. 14)				Schizophrenia (ref. 13)				Autism spectrum disorder ⁶⁻⁹				Intellectual disability ^{10,11}			
		NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)	
		<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut	<i>P</i>	# mut
FMRP targets (ALL)	784	0.0094	64	0.37	7	0.065	11	1	0	0.027	21	0.55	2	0.003	102	0.0003	26	2.00E-05	40	0.00068	10
FMRP targets not ARC complex	768	0.011	63	0.52	6	0.061	11	1	0	0.023	21	0.54	2	0.0046	100	0.00052	25	2.00E-05	35	0.0094	8
FMRP targets not NMDAR complex	753	0.016	61	0.67	5	0.055	11	1	0	0.062	19	0.84	1	0.0096	96	0.0004	25	2.00E-05	32	0.17	5
FMRP targets not ARC or NMDAR	745	0.014	61	0.67	5	0.053	11	1	0	0.059	19	0.84	1	0.012	95	0.00088	24	2.00E-05	31	0.35	4
FMRP targets excluding all PSD genes	615	0.02	51	0.68	4	0.037	10	1	0	0.12	15	0.77	1	0.013	80	0.0094	18	2.00E-05	29	0.22	4
ARC complex (ALL)	28	0.00048	6	0.005	2	1	0	1	0	1	0	1	0	0.22	3	0.22	1	2.00E-05	5	0.0015	2
ARC complex and FMRP target	16	0.46	1	0.068	1	1	0	1	0	1	0	1	0	0.26	2	0.14	1	2.00E-05	5	0.00084	2
ARC complex not FMRP targets	12	6.00E-05	5	0.045	1	1	0	1	0	1	0	1	0	0.47	1	1	0	1	0	1	0
NMDAR complex (ALL)	60	0.025	6	0.035	2	1	0	1	0	0.13	2	0.086	1	0.031	8	0.46	1	2.00E-05	8	2.00E-05	5
NMDAR complex and FMRP target	31	0.17	3	0.016	2	1	0	1	0	0.061	2	0.055	1	0.031	6	0.33	1	2.00E-05	8	2.00E-05	5
NMDAR complex not FMRP targets	29	0.04	3	1	0	1	0	1	0	1	0	1	0	0.36	2	1	0	1	0	1	0

Enrichment was tested using dnenrich (Supplementary Information). Columns are as in Table 3, and $P < 0.05$ are marked in bold. PSD, postsynaptic density.

Extended Data Table 4 | Brain expression biases of genes affected by *de novo* mutations

a

Mutations			Current study				Schizophrenia (ref. 14)				Schizophrenia (ref. 13)				Autism spectrum disorder ^{6,9}				Intellectual disability ^{10,11}			
Brain region	Expression bias?	Genes (N)	NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)	
			P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut
HPC	none	5373	0.72	106	0.1	19	0.72	14	0.52	3	0.41	33	0.48	5	0.36	186	0.54	30	0.95	25	0.8	6
	prenatal	6444	0.45	175	0.81	22	0.12	30	0.91	3	0.32	53	0.52	8	0.00028	332	0.021	63	0.14	57	0.21	16
	postnatal	7299	0.13	196	0.63	22	0.78	23	0.21	6	0.82	47	0.44	8	1	258	0.99	37	0.33	57	0.57	12
PFC	none	4997	0.36	104	0.41	14	0.99	7	0.72	2	0.9	23	0.94	2	0.92	149	0.89	22	0.89	24	0.71	6
	prenatal	6266	0.44	174	0.52	25	0.071	31	0.54	5	0.18	55	0.34	9	6.00E-05	333	0.00084	69	0.052	60	0.2	16
	postnatal	7853	0.34	200	0.59	24	0.37	29	0.5	5	0.59	54	0.35	9	0.97	294	0.99	39	0.68	55	0.69	12

b

Mutations			Current study				Schizophrenia (ref. 14)				Schizophrenia (ref. 13)				Autism spectrum disorder ^{6,9}				Intellectual disability ^{10,11}			
Brain expression?	Genes (N)		NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)	
			P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut
low	5851		0.89	118	0.97	11	0.48	19	0.17	5	0.31	40	0.44	6	0.99	185	1	22	1	23	1	2
high	9279		0.0058	264	0.016	40	0.45	34	0.57	6	0.12	74	0.34	11	2.00E-05	442	0.00018	86	6.00E-05	93	0.0007	26
prenatal	7962		0.33	225	0.39	32	0.74	29	0.97	3	0.2	68	0.65	9	0.00054	405	0.00024	83	0.35	67	0.21	19
postnatal	2393		0.17	64	0.54	7	0.68	7	0.74	1	0.96	10	0.65	2	0.95	79	0.88	11	0.71	15	0.89	2

a, Enrichment of *de novo* mutations (as calculated by dnenrich, see Supplementary Information) falling in genes with pre- or postnatal brain expression bias. Number and significance of overlap of mutations in schizophrenia, autism and intellectual disability in genes with no brain expression bias, or with a pre- or postnatal expression bias in the brain, based on Human Brain Transcriptome (HBT) data as used in ref. 13 (see Supplementary Information), for two brain regions; HPC, hippocampus; PFC, prefrontal cortex. Columns are as in Table 3, and *P* < 0.05 are marked in bold. **b**, Enrichment of *de novo* mutations (as calculated by dnenrich, see Supplementary Information) on the basis of RNA-seq-based brain expression and developmental trajectory. Number and significance of overlap of mutations in schizophrenia, autism and intellectual disability in genes not expressed in the brain, highly expressed in the brain, or with a pre- or postnatal expression bias in the brain (rows), based on BrainSpan RNA-seq data (see Supplementary Information). Columns are as in Table 3, and *P* < 0.05 are marked in bold.

Extended Data Table 5 | Comparison of genes hit by *de novo* mutations between this study and other disease studies and control individuals

Mutations (N)		Current study				Schizophrenia (ref. 14)				Schizophrenia (ref. 13)				Autism spectrum disorder ⁶⁻⁹				Intellectual disability ^{10,11}				Controls ^{7-10,13-14}			
		NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)		NS (434)		LoF (49)	
Gene set	Genes (N)	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut	P	# mut
Current study	NS (464)					0.16	6	0.03	3	0.85	4	0.29	1	0.016	55	0.0066	15	0.044	13	0.26	3	0.61	21	0.72	2
	LoF (63)			-		0.014	3	0.088	1	1	0	1	0	0.0023	14	0.00012	7	0.019	4	0.002	3	1	0	1	0
Schizophrenia (ref. 14)	NS (67)	0.22	6	0.021	3			-		0.31	2	0.16	1	0.47	7	0.11	3	0.67	1	1	0	0.48	4	1	0
	LoF (12)	0.051	3	0.11	1					0.21	1	1	0	0.15	3	0.21	1	0.21	1	1	0	1	0	1	0
Schizophrenia (ref. 13)	NS (136)	0.79	5	1	0	0.25	2	0.15	1					0.083	16	1	0	0.13	4	0.012	3	0.081	10	0.49	1
	LoF (20)	0.24	2	1	0	0.13	1	1	0			-		1	0	1	0	0.0026	3	4.00E-05	3	0.21	2	1	0
Autism spectrum disorder ⁶⁻⁹	NS (743)	0.14	45	0.023	9	0.49	6	0.13	3	0.32	14	1	0			-		6.00E-05	24	2.00E-05	12	0.26	38	0.64	4
	LoF (128)	0.015	11	0.00072	4	0.11	2	0.17	1	1	0	1	0					2.00E-05	8	2.00E-05	5	0.36	8	0.52	1
Intellectual disability ^{10,11}	NS (132)	0.032	9	0.031	1	0.56	1	0.14	1	0.14	2	0.01	1	2.00E-05	24	2.00E-05	7			-		0.37	7	0.46	1
	LoF (30)	0.27	1	0.019	1	1	0	1	0	0.046	1	0.0062	1	2.00E-05	15	2.00E-05	5					1	0	1	0
Controls ^{7-10,13-14}	NS (424)	0.59	21	1	0	0.41	4	1	0	0.15	9	0.26	2	0.062	41	0.31	8	0.48	7	1	0			-	
	LoF (49)	0.6	2	1	0	1	0	1	0	0.44	1	1	0	0.42	4	0.45	1	0.45	1	1	0				

Each set of columns gives the number of mutations (either nonsynonymous (NS) or loss-of-function (LoF)) and enrichment *P* value (as calculated by dnenrich, see Supplementary Information) in the set of genes hit by *de novo* mutations in the study listed in the corresponding row. For example, the first two rows detail the significance of the overlap of the mutations from other studies of disease hitting the genes hit by mutations in this study. Nominally significant *P* values (<0.05) are marked in bold. Disease sets and functional classes are as listed in Table 2.

Extended Data Table 6 | Mammalian conservation at *de novo* mutation sites and of genes hit by *de novo* mutations

(a) GERP score of NS Mutations

		Intellectual disability ^{10,11}	Autism spectrum disorder ⁶⁻⁹	Current study
	median (N)	4.89 (141)	4.72 (780)	4.48 (481)
Intellectual disability ^{10,11}	4.89 (141)	-	0.028	0.00053
Autism spectrum disorder ⁶⁻⁹	4.72 (780)	0.972	-	0.013
Current study	4.48 (481)	0.999	0.987	-

(b) GERP score of genes containing NS mutations

		Intellectual disability ^{10,11}	Autism spectrum disorder ⁶⁻⁹	Current study
	median (N)	4.75 (141)	4.27 (780)	4.2 (481)
Intellectual disability ^{10,11}	4.75 (141)	-	0.00015	0.000009
Autism spectrum disorder ⁶⁻⁹	4.27 (780)	0.9999	-	0.166
Current study	4.2 (481)	1.0000	0.834	-

(c) Variant term from joint linear model of variant and gene GERP scores

Comparison	Coefficient	P-value
Intellectual disability > Autism spectrum disorder	0.052	0.270
Intellectual disability > Current study	0.102	0.044
Autism spectrum disorder > Current study	0.039	0.079

Logistic regression model for (X > Y): type ~ gene_gerp + variant_gerp

a. Mann–Whitney rank test of the Genomic Evolutionary Rate Profiling (GERP) score (see Supplementary Information) distributions of nonsynonymous (NS) *de novo* mutations between pairs of phenotypes, with significant pairwise comparisons ($P < 0.05$) in bold. **b.** Mann–Whitney rank test of the median GERP scores of genes (see Supplementary Information) hit by nonsynonymous *de novo* mutations between pairs of phenotypes, with significant comparisons ($P < 0.05$) in bold. **c.** Linear modelling of variant GERP and per-gene GERP was employed to test whether the differences observed in **a** were driven by those observed in **b**. The coefficients and P value of the variant GERP score (from the joint linear models) are shown, where, for example, “intellectual disability > current study” indicates a test of whether the conservation at sites of *de novo* mutations in intellectual disability is greater than that of mutations in schizophrenia (from the current study), after correcting for the fact that the mutations in intellectual disability hit genes with greater overall conservation (**b**). $P < 0.05$ are marked in bold.

A polygenic burden of rare disruptive mutations in schizophrenia

Shaun M. Purcell^{1,2,3,4,5}, Jennifer L. Moran^{1*}, Menachem Fromer^{1,2,3,4*}, Douglas Ruderfer^{2,3*}, Nadia Solovieff⁴, Panos Roussos^{2,3}, Colm O'Dushlaine¹, Kimberly Chambert¹, Sarah E. Bergen^{1,6}, Anna Kähler⁶, Laramie Duncan^{1,4,5}, Eli Stahl^{2,3}, Giulio Genovese¹, Esperanza Fernández^{7,8}, Mark O. Collins⁹, Noboru H. Komiyama⁹, Jyoti S. Choudhary⁹, Patrik K. E. Magnusson⁶, Eric Banks⁵, Khalid Shakir⁵, Kiran Garimella⁵, Tim Fennell⁵, Mark DePristo⁵, Seth G. N. Grant¹⁰, Stephen J. Haggarty^{1,4,11}, Stacey Gabriel⁵, Edward M. Scolnick¹, Eric S. Lander⁵, Christina M. Hultman⁶, Patrick F. Sullivan¹², Steven A. McCarroll^{1,5,13} & Pamela Sklar^{2,3,14}

Schizophrenia is a common disease with a complex aetiology, probably involving multiple and heterogeneous genetic factors. Here, by analysing the exome sequences of 2,536 schizophrenia cases and 2,543 controls, we demonstrate a polygenic burden primarily arising from rare (less than 1 in 10,000), disruptive mutations distributed across many genes. Particularly enriched gene sets include the voltage-gated calcium ion channel and the signalling complex formed by the activity-regulated cytoskeleton-associated scaffold protein (ARC) of the postsynaptic density, sets previously implicated by genome-wide association and copy-number variation studies. Similar to reports in autism, targets of the fragile X mental retardation protein (FMRP, product of *FMR1*) are enriched for case mutations. No individual gene-based test achieves significance after correction for multiple testing and we do not detect any alleles of moderately low frequency (approximately 0.5 to 1 per cent) and moderately large effect. Taken together, these data suggest that population-based exome sequencing can discover risk alleles and complements established gene-mapping paradigms in neuropsychiatric disease.

Genetic studies of schizophrenia (MIM 181500) have demonstrated a substantial heritability^{1,2} that reflects common and rare alleles at many loci. Genome-wide association studies (GWAS) continue to uncover common single nucleotide polymorphisms (SNPs) at novel loci³. Rare or *de novo* genetic deletions and duplications (copy-number variants (CNVs)) have been firmly established, including risk variants at 22q11.2, 15q13.3 and 1q21.1 (refs 4, 5). One notable outcome of these large-scale, genome-wide investigations is the degree of polygenicity, consistent with thousands of genes and non-coding loci harbouring risk alleles^{3,6–9}.

Nonetheless, progress has been made in implicating biological systems and quantifying shared genetics among related psychiatric disorders (for example, refs 10, 11), such as identifying common variants in calcium ion channel genes affecting schizophrenia and bipolar disorder¹² and *de novo* CNVs affecting genes encoding members of the postsynaptic density (PSD) proteome¹³, in particular members of the neuronal ARC protein and *N*-methyl-D-aspartate receptor (NMDAR) postsynaptic signalling complexes.

Here we apply massively parallel short-read sequencing to assay a substantial portion of variation that previously was essentially invisible: rare coding point mutations (single nucleotide variants (SNVs)) and small insertions and deletions (indels). Although previous schizophrenia studies have applied sequencing, the results have been inconclusive, reflecting limited sample sizes or a focus on small numbers of candidate genes^{14–17}. Exome-sequencing studies of *de novo* mutations published to date have neither demonstrated an increased rate in schizophrenia, nor conclusively implicated individual genes^{18,19}, although some data suggest a link with particular classes of gene, such as those with

higher brain expression in early fetal life¹⁹. *De novo* studies in intellectual disability^{20,21} and autism^{22–25} have, however, made considerable progress in identifying large-effect alleles and the underlying gene networks.

In this study, we sought to identify the alleles, genes or gene networks that harbour rare coding variants of moderate or large effect on risk for schizophrenia by exome sequencing 5,079 individuals, selected from a Swedish sample of more than 11,000 individuals. Previous analyses of the full sample (Supplementary Information section 1) have demonstrated an enriched burden of rare CNVs and a polygenic common variant component³. We generated high-coverage exome sequence to ensure sufficient sensitivity to detect and genotype alleles observed in only one heterozygous individual (singletons, implying an allele frequency of ~1 in 10,000, although the true population frequency will typically be rarer).

The high baseline rate of rare, neutral mutations makes it difficult to detect rare alleles that increase risk for common diseases²⁶. Although power can be increased by jointly testing groups of variants in a gene²⁷, association testing across all genes is likely to be under-powered at current sample sizes. Indeed, a recent application of population-based exome sequencing in autism did not identify genes²⁸, despite moderately large sample size and the success of the *de novo* paradigm. Furthermore, many confirmed results from candidate-gene sequencing studies of nonpsychiatric disease still fall short of exome-wide significance²⁹.

We therefore adopted a top-down strategy in which we studied a large set of genes with a higher likelihood of having a role in schizophrenia, on the basis of existing genetic evidence (Supplementary Information section 7). We focused on ~2,500 genes implicated by unbiased,

¹Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁴Analytic and Translational Genetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden. ⁷Center for Human Genetics, KU Leuven, 3000 Leuven, Belgium. ⁸VIB Center for Biology of Disease, 3000 Leuven, Belgium. ⁹Proteomic Mass Spectrometry, The Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. ¹⁰Genes to Cognition Programme, Centre for Clinical Brain Sciences and Centre for Neuroregeneration, The University of Edinburgh, Edinburgh EH16 4SB, UK. ¹¹Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹²Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA. ¹³Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁴Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

*These authors contributed equally to this work.

large-scale genome-wide screens, including GWAS, CNV and *de novo* SNV studies, testing for enrichment of rare alleles in cases. To prioritize individual genes, we characterized emerging signals with respect to the genes and frequency and type of mutations. We coordinated analysis with an independent trio exome-sequencing study (Fromer *et al.*³⁰, this issue) and note key points of convergence below.

After alignment and variant calling of all samples jointly, we removed 11 subjects with low-quality data along with likely spurious sites and genotypes (Supplementary Information sections 2 and 3). Per individual, 93% of targeted bases were covered at ≥ 10 -fold (81% at ≥ 30 -fold). The final data set comprised 2,536 cases and 2,543 controls (Extended Data Table 1a and Extended Data Fig. 1a). Cases and controls had similar technical sequencing metrics, including total coverage, proportion of deeply covered targets, and overall proportion of non-reference alleles (Extended Data Table 1b). We observed 635,944 coding and splice-site passing variants of which 56% were singletons. Using Sanger sequencing and Exome Chip data on these samples, we determined high specificity and sensitivity for singletons (Supplementary Information section 3).

We annotated variants with respect to RefSeq and combined five *in silico* algorithms to predict missense deleteriousness (Extended Data Table 1c and Supplementary Information section 4). As expected, allelic types more likely to affect protein function showed greater constraint: 69% of nonsense variants were singletons, compared to 58% of missense and 51% of silent variants. Primary analyses tested (1) disruptive variants (nonsense, essential splice site and frameshifts, $n = 15,972$ alleles with minor allele frequency (MAF) $< 0.1\%$); (2) disruptive plus missense variants predicted to be damaging by all five algorithms ($n = 50,369$); and (3) disruptive plus missense variants predicted to be damaging by at least one algorithm ($n = 233,575$). These groups are labelled disruptive, NS_{strict} and NS_{broad}, in which NS indicates nonsynonymous. We also stratified most analyses by allele frequency: (1) singletons; (2) up to 0.1% (ten or fewer minor alleles); and (3) up to 0.5% (50 or fewer minor alleles). In the main gene set analyses, we empirically corrected for multiple testing over the nine combinations of these factors (Supplementary Information section 7).

The most significant SNV or indel association ($P = 5 \times 10^{-8}$) was for a common missense allele in *CCHCR1*, in the major histocompatibility complex (MHC), a known risk locus; this top SNP was in linkage disequilibrium with many other schizophrenia-associated SNPs in the MHC. All $P < 10^{-5}$ variants were for either common alleles or a

few instances of likely aberrant variants that had escaped earlier filtering (Supplementary Information section 5). We performed two series of gene-based tests: a one-sided burden test of an increased rare allele rate in cases, and the SNP-set (sequence) kernel association test (SKAT²⁷), which allows for risk and protective effects. For both tests, the distribution of gene-based statistics broadly followed a global null (Extended Data Fig. 1b).

Considering only disruptive variants, the genic test yielding the lowest nominal P value was for *KYNU* (kynureninase), showing ten variants in cases and zero in controls (Extended Data Table 2 and Supplementary Table 1); one novel nonsense mutation at chr2:143713804 (g.468T>A; p.Y156*) was observed in seven cases and not present in either the Exome Variant Server (<http://evs.gs.washington.edu/EVS/>) or 1000 Genomes Project (<http://www.1000genomes.org/>). Although previous studies have suggested links between the kynurenine pathway and schizophrenia³¹, our P value of 1.7×10^{-3} does not withstand correction for multiple testing, even if considering only the 246 genes with ≥ 10 rare disruptive mutations capable of achieving a nominally significant result.

A polygenic burden of rare coding variants

We evaluated a polygenic burden of rare coding variants in cases, first selecting 2,546 genes ($\sim 10\%$ of the exome) on the basis of previous genetic studies that we proposed to be enriched for schizophrenia-associated mutations (Supplementary Information section 6). Sources included genome-wide CNV studies^{5,13}, GWAS^{3,12,32} and exome sequencing of *de novo* mutations^{18,19,30}. In our sample, these genes had a significantly higher rate of rare (MAF $< 0.1\%$), disruptive mutations in cases compared to controls ($P = 10^{-4}$ for 1,547 versus 1,383 mutations). The enrichment was unlikely to represent technical or ancestry-related artefact because the P values controlled for potential differences in exome-wide burden in cases and controls, and because we observed no differences exome wide ($P = 0.24$). Furthermore, enrichment P values were empirically derived by permuting phenotypes within subgroups of cases and controls, matched on exome-wide identity-by-state, experimental batch and sex; the above result withstood correction for multiple testing (Table 1). We observed similar results for rarer (singletons, $P = 8 \times 10^{-4}$) and more frequent (MAF $< 0.5\%$, $P = 2 \times 10^{-4}$) alleles. We also observed case enrichment for the strictly defined set of damaging mutations (NS_{strict}, $P = 1.5 \times 10^{-3}$) but not the broader set (NS_{broad}, $P = 0.13$).

Table 1 | Gene set analysis of primary schizophrenia candidate gene sets

Variant type	Gene set/subset	<i>n</i> genes	Singletons	MAF $< 0.1\%$	MAF $< 0.5\%$
Disruptive	Primary	2,546	0.0008	0.0001	0.0002
NS _{strict}			0.0059	0.0015	0.0110
NS _{broad}			0.0986	0.1295	0.1126
Disruptive	SCZ <i>de novo</i> genes				
	Exome sequencing (disruptive) ^{18,19,30}	87	0.0319	0.0007	0.0003
	Exome sequencing (nonsyn.) ^{18,19,30}	611	0.0053	0.0011	0.0055
	Copy number variants				
	<i>de novo</i> CNV genes ¹³	234	0.0234	0.0039	0.0124
	SCZ-associated CNV genes ⁵	345	0.3308	0.4596	0.4376
	GWAS				
	Voltage-gated calcium channel genes ¹²	26	0.0019	0.0214	0.0212
	Common SNPs ($P < 10^{-4}$ intervals) ³	479	0.1794	0.0368	0.0037
	miR-137 targets ³²	446	0.6573	0.5609	0.4747
	Synaptic genes				
	PSD (human core) ¹³	685	0.0808	0.1154	0.1256
	ARC ¹³	28	0.0016	0.0014	0.0014
	NMDAR network ¹³	61	0.0158	0.0251	0.0252
	PSD-95 ¹³	65	0.0017	0.0009	0.0010
	mGluR5 ¹³	39	0.1327	0.0900	0.0902

Enrichment test empirical P values for rare (singleton; MAF $< 0.1\%$; MAF $< 0.5\%$) variants from disruptive, NS_{strict} and NS_{broad} sets. P values represent the relative case enrichment compared to average exome-wide case/control difference. Bolded values are significant at $P_{corrected} < 0.05$. Initial comparison corrects (based on the empirical distribution of minimum P values) for the nine correlated tests (top panel). The bottom panel focuses on the 12 subsets of the primary gene set, for disruptive variants only as they showed the greatest enrichment for the entire primary set. Again, bold values are significant after correcting for the 36 tests performed.

SCZ, schizophrenia.

This enrichment suggests a polygenic burden of rare variants. Although not so marked as to be detectable at the exome-wide level given the sample size, it is relatively concentrated in genes that were found to be associated with schizophrenia by other methods. The mean allelic effect was not large: in the primary comparison, the odds ratio was 1.12 (1.04–1.20, 95% confidence interval) for each MAF < 0.1% disruptive mutation; 46% of cases carried one or more allele in this primary set (0.62 per case) compared to 41% of controls (0.55 per control). At two extremes, the modest mean effect could represent either that a subset of mutations are fully penetrant or that every allele is associated but increases risk by only 12%, similar to common alleles from GWAS. To extract subsets of potentially stronger-effect alleles, we individually tested the constituent gene sources (Table 1 and Extended Data Fig. 1c), focusing on disruptive variants as they showed the strongest omnibus enrichment. For disruptive mutations, eight out of 12 sets were nominally significant ($P < 0.05$), indicating that the initial observation was not driven by a single category.

ARC, PSD-95 and calcium ion channel genes

Three of the smaller significantly enriched sets (the ARC and PSD-95 (encoded by *DLG4*) complexes and calcium ion channel genes) had odds ratios > 5. We observed enrichment ($P = 1.6 \times 10^{-3}$) of disruptive mutations among the 28 ARC complex genes: nine mutations in nine genes (all singletons) in cases and zero in controls, yielding an odds ratio of 19.2 (2.4–2,471, 95% confidence interval; Extended Data Table 2). Along with the NMDAR gene set (also significantly enriched), ARC genes largely accounted for the overall PSD enrichment ($P = 4 \times 10^{-8}$) in ref. 13, in which four ARC genes had one or more *de novo* CNVs. Of note, in an independent exome-sequencing study in trios, Fromer *et al.*³⁰ found that the ARC gene set was enriched ($P = 5 \times 10^{-4}$) for nonsynonymous *de novo* SNVs and indels, with four genes harbouring six mutations (Extended Data Table 7). The other PSD gene set with strong enrichment ($P = 9 \times 10^{-4}$; odds ratio = 5.1, 1.8–19.2, 95% confidence interval) was the PSD-95 complex, which contains 65 genes and overlaps with ARC. PSD genes are very highly conserved and have critical roles in excitatory neural signalling components, as well as dendrite and spine plasticity. Further categorization of neuronal genes on the basis of subcellular localization¹³ (Extended Data Table 3a) or associated mouse and human phenotypes³³ did not yield further enrichment.

The other subset yielding a large odds ratio of 8.4 (2.03–77, 95% confidence interval) was the 26 voltage-gated calcium ion channel genes (12 cases, one control; disruptive singletons, $P = 2 \times 10^{-3}$, although the effect is attenuated when including recurrent alleles: 15/8 cases/controls, $P = 0.021$, see Extended Data Table 2). The singleton enrichment was predominantly driven by the pore-forming α_1 and auxiliary $\alpha_2\delta$ subunits; of the α_1 subunits, the Ca_v1/L -type genes carried the most case mutations, including two in *CACNA1C*, a gene implicated by GWAS of bipolar disorder and schizophrenia^{3,10}. Calcium signalling is involved in many cell functions including the regulation of gene expression³⁴ and is critical for modulating synaptic plasticity³⁵. In a secondary analysis of proteins found in the nano-environment of the calcium channel³⁶, we observed independent enrichment for other ion channel transporters (Supplementary Table 1), odds ratio 9.1 (2.2–83) for singletons ($P = 1 \times 10^{-3}$; 13/1 disruptive alleles).

Convergence with *de novo* studies

A line of convergence across studies was that genes carrying nonsynonymous *de novo* mutations^{18,19,30} were enriched for rare disruptive mutations in cases ($P = 1 \times 10^{-3}$; Table 1 and Extended Data Table 6a, b). We observed a similar result for the smaller class of genes carrying disruptive *de novo* mutations ($P = 7 \times 10^{-4}$, from 47 genes in our study); these genes included *UFL1* (5/0 disruptive mutations, $P = 0.03$; 7/0 $\text{NS}_{\text{strict}}$, $P = 0.008$), *SYNGAP1* (4/0 $\text{NS}_{\text{strict}}$, $P = 0.04$) and *SZT2* (18/9 $\text{NS}_{\text{strict}}$, $P = 0.049$). *SYNGAP1* (synaptic Ras GTPase activating protein 1) is a component of the NMDAR PSD complex³⁷ and mutations in this gene are known to cause intellectual disability and autism³⁸.

Genes under previously associated CNV regions did not show significant enrichment of rare disruptive mutations, although there was an enrichment of $\text{NS}_{\text{strict}}$ mutations ($P = 0.0044$; Extended Data Table 4). Of the 11 CNV regions, only the 3q29 locus, which contains multiple genes including *DLG1* (ref. 4), was significant ($P = 0.0006$) and withstood correction for multiple testing.

Autism/intellectual disability genes and FMRP targets

We next tested, as a single set, the 2,507 genes representing autism and intellectual disability candidates (Supplementary Information section 6), which yielded only nominal significance ($P < 0.05$) for disruptive and $\text{NS}_{\text{strict}}$ variants and no test survived correction for multiple testing (Table 2). Considering the 12 constituent sets, genes from autism *de novo*

Table 2 | Gene set analysis of secondary autism/intellectual disability candidate gene sets

Variant type	Gene set/subset	<i>n</i> genes	Singletons	MAF < 0.1%	MAF < 0.5%
Disruptive	Autism/ID	2,507	0.029	0.043	0.049
$\text{NS}_{\text{strict}}$			0.052	0.008	0.013
NS_{broad}			0.532	0.619	0.287
		<i>n</i> genes	Min. $P^{\text{corrected}}$ (for $9 \times 12 = 108$ tests)		
Disruptive, $\text{NS}_{\text{strict}}$ and NS_{broad}	<i>De novo</i> genes (exome sequencing)				
	Autism (disruptive) ^{22–25}	128		1.000	
	Autism (nonsyn.) ^{22–25}	743		1.000	
	ID (disruptive) ^{20,21}	30		0.070	
	ID (nonsyn.) ^{20,21}	132		0.995	
	Neurodevelopmental candidates				
	ASD candidates ³⁹	112		1.000	
	ID candidates ³⁹	196		1.000	
	Autism PPI networks				
	CHD8 network ²⁴	6		1.000	
	49-gene network ²⁴	49		1.000	
	74-gene network ²⁴	74		1.000	
	Fragile X mental retardation protein targets				
	Darnell targets ⁴⁰	788		0.010	
	Ascano targets ⁴²	939		0.997	
	Ascano FMRP/autism overlap ⁴²	93		0.993	

Enrichment test empirical *P* values for the secondary (autism/intellectual disability) gene set. As in Table 1, the top panel shows uncorrected *P* values; tests significant after multiple test correction are in bold (that is, all $P^{\text{corrected}} > 0.05$). Because no class of variant is significant after multiple test correction for the omnibus test (top panel), we applied and corrected for all 108 tests (nine conditions by 12 subsets) in the bottom panel. The single-category FMRP targets (Darnell *et al.*⁴⁰) mainly reflect disruptive and $\text{NS}_{\text{strict}}$ singleton enrichment. ASD, autism spectrum disorder; ID, intellectual disability; PPI, protein–protein interaction.

studies showed no enrichment (Extended Data Fig. 1c), despite greater sample size and number of disruptive *de novo* mutations. There was no evidence for autism or intellectual disability genes curated from the literature³⁹ or for genes in the protein–protein-interaction-derived subnetworks built around autism *de novo* mutations²⁴.

The nominal omnibus signals arose largely from the Darnell *et al.* list of FMRP targets⁴⁰. FMRP is encoded by the gene *FMR1* (the locus of the Mendelian fragile X syndrome repeat mutation) and is an RNA-binding protein that regulates translation and is needed at synapses for normal glutamate receptor signalling and neurogenesis⁴¹. Targets of FMRP are enriched for *de novo* mutations in autism^{22,40,42}; here we find significant enrichment of disruptive singletons ($P = 1.4 \times 10^{-3}$; 289/223 case/control count; odds ratio = 1.3). These FMRP targets overlap with PSD genes (Extended Data Table 3b), although were still enriched independently (Supplementary Information section 6). In addition, these genes were enriched in GWAS of this sample ($P < 10^{-3}$, Supplementary Information section 9). Whereas the Darnell list is derived from mouse brain, a second recently reported FMRP target list⁴² was generated from cultured human embryonic kidney cells, using a different experimental approach (Supplementary Information section 6). This list has relatively little overlap with Darnell targets and, in contrast to the Darnell list, does not show any enrichment for rare case mutations, for GWAS loci, or comparable overlap with PSD genes (Extended Data Table 3b).

Our results are perhaps surprising: unlike Fromer *et al.*³⁰, we did not observe direct evidence for overlap at the individual gene level with autism and intellectual disability, despite CNV studies showing pleiotropic effects of individual loci. Nonetheless, at the broader level of gene sets, all three disorders showed enrichment for FMRP targets; autism and intellectual disability *de novo* mutations also showed strong enrichment in several PSD complexes enriched in our study, including NMDAR, PSD-95 and (for intellectual disability) ARC. At the least, our results suggest that any overlap is far from complete, although more refined analyses in larger samples will be needed before a clearer picture can emerge of which genes and pathways are shared and which are specific to one disease.

Characterizing enrichment by variant type

To further characterize the observed enrichment with respect to mutational function and frequency, we created a single ‘composite’ set of 1,796 genes comprising all members of the most prominently enriched sets (Supplementary Table 2). Rare disruptive mutations in this set were present in 990 cases and 877 controls (for singletons, 645–530). Cases carrying rare disruptive mutations did not appear to be phenotypically or clinically unusual in terms of sex, ancestry, history of drug abuse, general medical conditions plausibly aetiologically related to psychosis, or epilepsy, although they did have a higher rate of admissions noting comorbid intellectual disability compared to other cases ($P = 0.009$; Extended Data Table 2b).

Figure 1 shows composite set enrichment across a range of conditions. As this set merges other sets showing enrichment, it necessarily shows enrichment; it was not, however, due to confounding effects of ancestry, sex or experimental wave (Supplementary Information section 8). It was primarily driven by singleton nonsense mutations across a large number of genes, as it was removed or greatly attenuated when either singleton or nonsense mutations were excluded. Considered alone, neither splice-site, frameshift, missense, silent or noncoding mutations showed enrichment at $P < 0.01$. Different ways of defining damaging missense mutations did not substantively affect results. Considering only nonsynonymous coding variants present on Exome Chip, we did not observe enrichment. Rather, enrichment mainly reflected novel variants (Extended Data Table 5b), which is expected as most rare variants in our study are novel. We also took an alternative approach, whereby instead of filtering variants on frequency, we excluded genes with any control disruptive variants before calculating the burden of case alleles; the composite set was still highly enriched (‘case-unique’ in

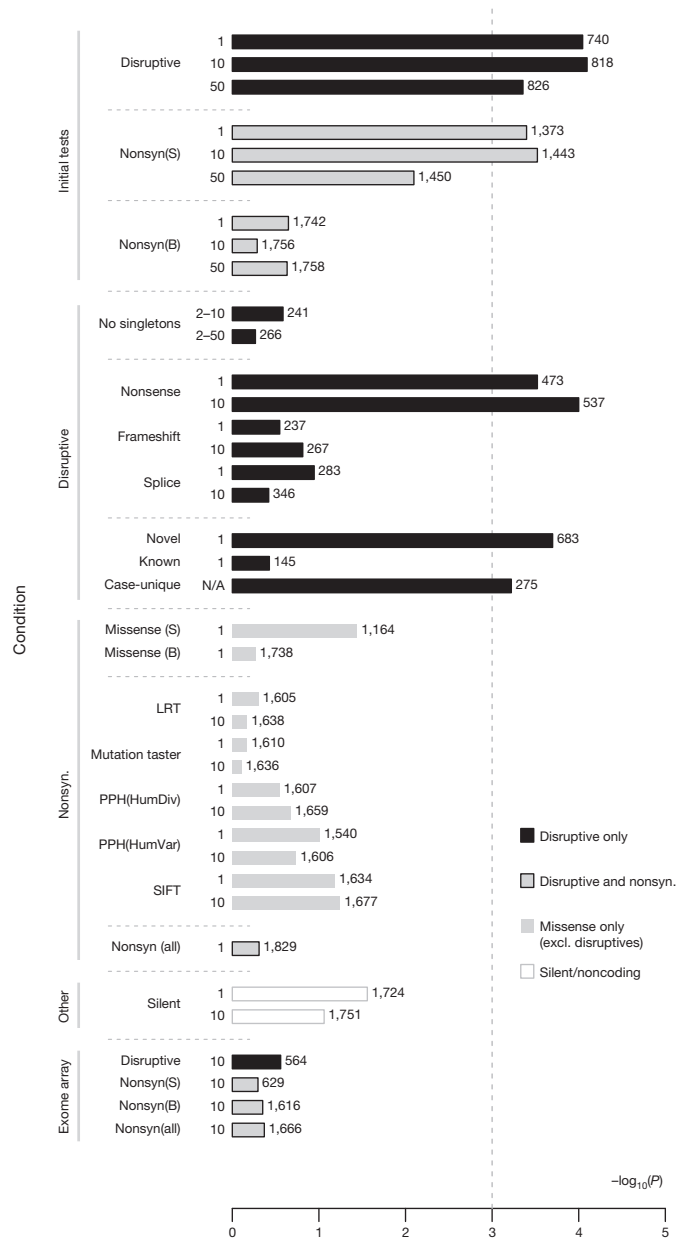


Figure 1 | Composite set gene set analysis, stratified by mutation type. Statistical significance (x axis) for the composite gene set stratified by type and frequency of mutation and other variables. Numbers to the right of each bar represent the number of genes with at least one mutation in that category for the composite set. (S) represents strictly defined damaging missenses; (B) represents the broadly defined group. Nonsyn (all) represents all nonsynonymous mutations. Numbers to the left of the bars (1, 10, 50) represent the minor allele count threshold (i.e. 1 indicates a singleton-only analysis); here the ranges 2–10 and 2–50 represent analyses that excluded singletons; N/A indicates that no allele-wise threshold was used. The source of deleteriousness prediction algorithms (LRT, MutationTaster, PPH and SIFT) is described in the Supplementary Information. For the exome array contrasts, Exome Chip sites were tested using the exome sequence calls.

Fig. 1; see Extended Data Table 5b and Supplementary Information section 7). Finally, the enrichment could not be attributed to only a small number of variants or genes (Extended Data Fig. 2a).

These findings do not preclude potentially important effects from other classes of rare variation in specific genes or other gene sets, although exploratory analyses of generic gene sets (for example, based on Gene Ontology terms) did not unambiguously identify novel signals after correction for multiple testing (Supplementary Information section 7).

We found preferential enrichment in genes with high brain expression, but not for genes with a prenatally biased developmental trajectory (Extended Data Fig. 3). In fact, greater enrichment came from postnatally biased genes. Finally, although greatly attenuated compared to disruptive mutations, other categories displayed nominal ($0.01 < P < 0.05$) enrichment in Fig. 1 and strictly defined damaging missense mutations alone showed enrichment for ARC and NMDAR gene sets (32/15 for ARC, $P = 0.007$; Extended Data Tables 5a and 7). Although rare coding alleles other than ultra-rare nonsense mutations will undoubtedly contribute to risk, it will probably prove harder still to elucidate such effects.

Rare variants, CNVs and common GWAS variants

We quantified the relative impact of common SNPs (indexed by a genome-wide polygene score from independent GWAS samples³²), rare CNVs (the burden of genic deletions) and disruptive mutations in the composite set. Considering the same 5,079 individuals, all three classes of variation were uncorrelated and significantly, independently and additively enriched in cases compared to controls. From logistic regression, the relative effect sizes (reduction in model R^2) were 5.7%, 0.2% and 0.4% for GWAS, rare CNV and rare coding variants, respectively (Supplementary Information section 8). Although not a complete assessment, it indicates that for the current sets of identifiably enriched alleles, common GWAS variants account for an order-of-magnitude more heritability than this set of rare variants does. However, these estimates will be diluted to varying degrees, owing to associated variants being included. As a consequence of this, and also the fact that true risk variants outside of composite set genes were not considered here, this estimate represents a conservative lower bound on the contribution of rare coding variation.

Discussion

We have demonstrated a polygenic burden that increases risk for schizophrenia, primarily comprising many ultra-rare nonsense mutations distributed across many genes. Implicating individual genes remains challenging, as genes that contributed to the highest-ranked sets typically had unremarkable P values, often around 0.5 with the gene containing only one or two rare mutations. Nonetheless, we were able to detect several small and highly enriched sets, notably of genes related to calcium channels and the postsynaptic ARC complex. Across these ~50 genes, ~1% of cases carried a rare disruptive mutation likely to have a considerable impact on risk. However, reported effect sizes will have a tendency to over-estimate true population values (Supplementary Information section 5).

We add to previous work that has implicated disruption of synaptic processes in schizophrenia¹³. The PSD is comprised of supramolecular multiprotein complexes that detect and discriminate patterns of neuronal activity and regulate plasticity processes responsible for learning⁴³. Members of the membrane-associated guanylate kinase (MAGUK) family of scaffold proteins, such as PSD-95, have a key role in assembling ~2 MDa complexes comprising calcium channels, including the glutamate-gated NMDAR, voltage-gated calcium channels and ARC^{36,44}. The genetic disruption of MAGUKs and their associated components result in specific cognitive impairments in mice and humans⁴⁵. One possibility is that the genetic risk identified here reflects altered tuning in calcium-dependent signalling cascades, triggered by NMDAR⁴⁶ and L-type calcium channels⁴⁷, mediated by postsynaptic MAGUK signalling complexes driving ARC synthesis.

Although we cannot yet use rare mutations to partition patients into more homogeneous clinical subgroups, this will remain a central goal for future sequencing studies. The few population-based common-disease exome-sequencing studies published to date, in psychiatric (for example, ref. 28) and non-psychiatric (for example, ref. 48) diseases, have not been successful in finding individual genes showing significant enrichment. Our study yields similar findings for individual genes, but yields positive results when considering gene sets. These current findings are likely to foreshadow the definitive identification of individual

genes in larger cohorts, following the trajectory of GWAS and other genetic studies of complex disease.

METHODS SUMMARY

Sample ascertainment. Cases with schizophrenia were identified through the Swedish Hospital Discharge Register³. Case inclusion criteria: ≥ 2 hospitalizations with a discharge diagnosis of schizophrenia, both parents born in Scandinavia, age ≥ 18 years. Case exclusion criteria: hospital register diagnosis of any disorder mitigating a confident diagnosis of schizophrenia. Controls were randomly selected from Swedish population registers. Control inclusion criteria: never hospitalized for schizophrenia or bipolar disorder, both parents born in Scandinavia, age ≥ 18 years. All subjects provided informed consent; institutional human subject committees approved the research.

Sequencing. The samples (2,536 cases, 2,543 controls) were sequenced using either the Agilent SureSelect Human All Exon Kit (29 Mb, $n = 132$) or the Agilent SureSelect Human All Exon v.2 Kit (33 Mb). Sequencing was performed by IlluminaGAII or Illumina HiSeq2000. Sequence data were aligned and variants called by the Picard (<http://picard.sourceforge.net/>)/BWA⁴⁹/GATK⁵⁰ pipeline. Validation of selected variants used Sanger sequencing. On the basis of validation and Exome Chip data, we estimated high sensitivity and specificity of singleton calls. BAM and VCF files are available in the dbGaP study phs000473.v1 (<http://research.mssm.edu/statgen/sweden/>).

Analysis. We used PLINK/SEQ (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000473.v1.p1) to annotate variants according to RefSeq gene transcripts (UCSC Genome Browser, <http://genome.ucsc.edu/>). Single-site association used Fisher's exact test; primary gene-based association used a burden test and the sequence kernel association test²⁷. Analyses controlled for ancestry and quality control metrics. Gene sets were evaluated on the empirical distribution of the sum of individual gene burden statistics, and incorporated an empirical correction for multiple testing. Odds ratios with 95% confidence intervals used penalized maximum likelihood (Firth's method) for low cell counts. See Supplementary Information for further details. Summary results are posted at <http://research.mssm.edu/statgen/sweden/>.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 24 December 2013.

Published online 22 January 2014.

- Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
- Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
- Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45**, 1150–1159 (2013).
- Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* **168**, 302–316 (2011).
- Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev. Genet.* **13**, 537–551 (2012).
- Stone, J. L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
- Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genet.* **44**, 247–250 (2012).
- Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241 (2012).
- Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nature Genet.* **43**, 977–983 (2011).
- Moreno-De-Luca, D. *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.* **87**, 618–630 (2010).
- Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
- Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
- Need, A. C. *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* **91**, 303–312 (2012).

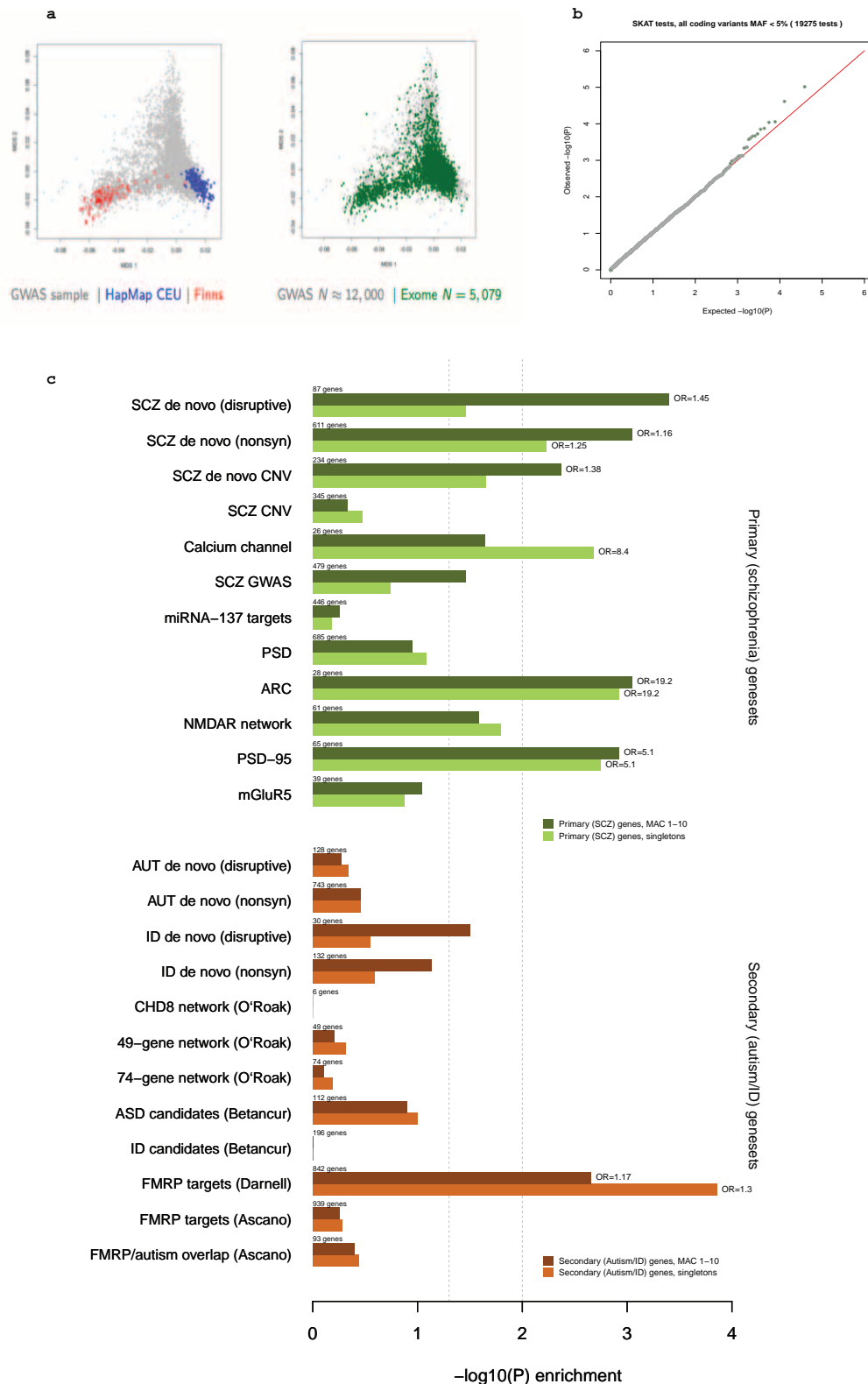
15. Crowley, J. J. *et al.* Deep resequencing and association analysis of schizophrenia candidate genes. *Mol. Psychiatry* **18**, 138–140 (2013).
16. Takata, A. *et al.* A population-specific uncommon variant in *GRIN3A* associated with schizophrenia. *Biol. Psychiatry* **73**, 532–539 (2013).
17. Timms, A. E. *et al.* Support for the *N*-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA Psychiatry* **3**, 1–9 (2013).
18. Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genet.* **43**, 860–863 (2011).
19. Xu, B. *et al.* *De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature Genet.* **44**, 1365–1369 (2012).
20. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
21. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
22. Iossifov, I. *et al.* *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
23. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
24. O’Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
25. Sanders, S. J. *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
26. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740–743 (2012).
27. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
28. Liu, L. *et al.* Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet.* **9**, e1003443 (2013).
29. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genet.* **44**, 623–630 (2012).
30. Fromer, M. *et al.* *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* <http://dx.doi.org/10.1038/nature12929> (this issue).
31. Linderholm, K. R. *et al.* Increased levels of kynurenine and kynurenine acid in the CSF of patients with schizophrenia. *Schizophr. Bull.* **38**, 426–432 (2012).
32. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genet.* **43**, 969–976 (2011).
33. Bayés, A. *et al.* Characterisation of the proteome, diseases and evolution of the human postsynaptic density. *Nature Neurosci.* **14**, 19–21 (2011).
34. Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**, 933–936 (1998).
35. Yasuda, R., Sabatini, B. L. & Svoboda, K. Plasticity of calcium channels in dendritic spines. *Nature Neurosci.* **6**, 948–955 (2003).
36. Müller, C. S. *et al.* Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. *Proc. Natl Acad. Sci. USA* **107**, 14950–14957 (2010).
37. Komiyama, N. H. *et al.* SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J. Neurosci.* **22**, 9721–9732 (2002).
38. Berryer, M. H. *et al.* Mutations in *SYNGAP1* cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. *Hum. Mutat.* **34**, 385–394 (2013).
39. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
40. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
41. Callan, M. A. & Zarnescu, D. C. Heads-up: new roles for the fragile X mental retardation protein in neural stem and progenitor cells. *Genesis* **49**, 424–440 (2011).
42. Ascano, M., Jr *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
43. Migaud, M. *et al.* Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. *Nature* **396**, 433–439 (1998).
44. Fernández, E. *et al.* Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol. Syst. Biol.* **5**, 269 (2009).
45. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neurosci.* **16**, 16–24 (2013).
46. Steward, O. & Worley, P. F. Selective targeting of newly synthesized *Arc* mRNA to active synapses requires NMDA receptor activation. *Neuron* **30**, 227–240 (2001).
47. Waltereit, R. *et al.* Arg3.1/*Arc* mRNA induction by Ca^{2+} and cAMP requires protein kinase A and mitogen-activated protein kinase/extracellular regulated kinase activation. *J. Neurosci.* **21**, 5484–5493 (2001).
48. Albrechtsen, A. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298–310 (2013).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler Transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful for the participation of all subjects contributing to this research, and to the collection team that worked to recruit them: E. Flordal-Thelander, A.-B. Holmgren, M. Hallin, M. Lundin, A.-K. Sundberg, C. Pettersson, R. Satgunathan-Dawoud, S. Hasselund, M. Rådstrom, B. Ohlander, L. Nyrén and I. Kizling. We acknowledge funding support from National Institutes of Health (NIH)/National Institute of Mental Health (NIMH) ARRA Grand Opportunity grant NIMH RC2 MH089905 (S.M.P., P.S.), the Sylvan Herman Foundation, the Stanley Center for Psychiatric Research, the Stanley Medical Research Institute, NIH/National Human Genome Research Institute (NHGRI) grant U54HG003067 (E.S.L.), NIH/NIMH grant R01 MH095088 (S.J.H.), NIH/NIMH grant R01 MH091115 (S.J.H.), the Tau Consortium (S.J.H.), NIH/NIMH grant R01 MH099126 (S.M.P.), NIH/NHGRI grant R01 HG005827 (S.M.P.), NIH/NIMH grant R01 MH077139 (P.F.S.), NIH/NIMH grant R01 MH095034 (P.S.), NIH/NIMH grant T32 MH017119 (L.D.), the Friedman Brain Institute at Mount Sinai School of Medicine, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, an ALF grant from Swedish County Council, the Söderström Königska Foundation, the Netherlands Scientific Organization (NWO 645-000-003), the Wellcome Trust, Genes to Cognition Program, The Medical Research Council and European Union projects GENCODYN no. 241995, EUROSPIN no. 242498 and SYNSYS no. 242167 (E.F., M.O.C., N.H.K., J.S.C., S.G.N.G.). Work at the Icahn School of Medicine at Mount Sinai was also supported by the Institute for Genomics and Multiscale Biology (including computational resources and staff expertise provided by the Department of Scientific Computing). The funders had no role in study design, execution, analysis or manuscript preparation.

Author Contributions Project leadership: S.M.P., J.L.M., P.F.S., S.A.M., C.M.H., P.S. Sample collection and phenotyping: A.K., P.K.E.M., P.F.S., K.C., J.L.M., C.M.H. Sample processing and data management: K.C., D.R., M.F., J.L.M. Sequencing and variant calling: M.d.P., E.B., K.S., K.G., T.F., S.G. Primary statistical analysis: D.R., M.F., S.M.P. Additional analyses: L.D., E.S., G.G., S.J.H., N.S., P.R., C.O’D., S.E.B. Determination of synaptic gene sets: S.J.H., E.F., M.O.C., N.H.K., J.S.C., S.G.N.G. Interpretation of main findings: S.M.P., E.M.S., E.S.L., S.J.H., M.F., P.F.S., S.A.M., P.S. Primary drafting of manuscript: S.M.P., P.F.S., S.A.M., E.S.L., P.S. Production and approval of the final manuscript: all authors.

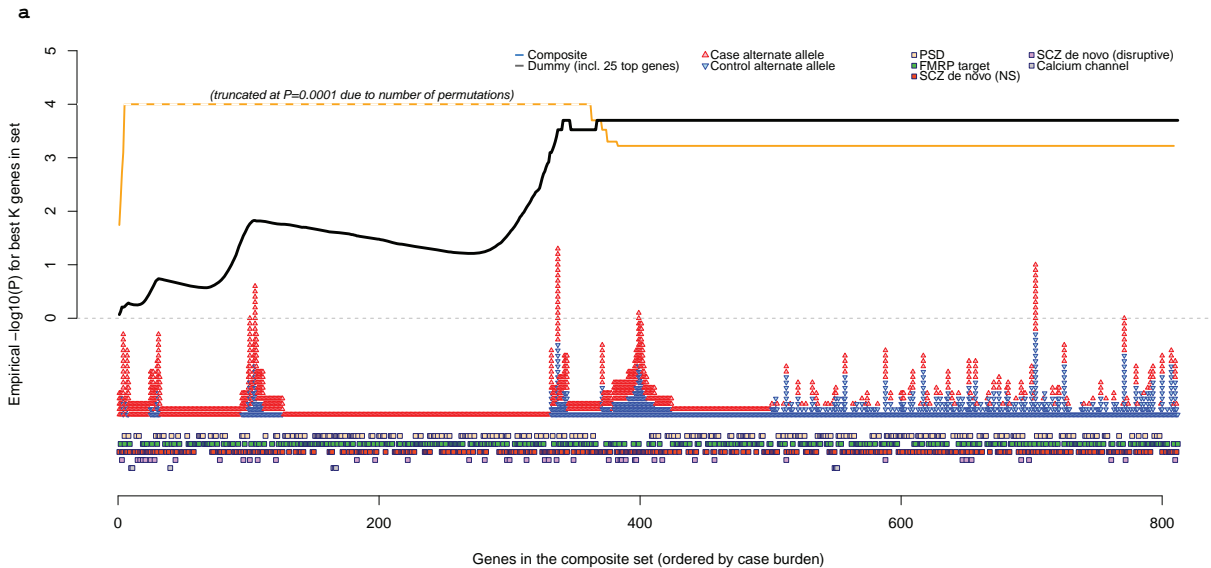
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.P. (shaun@broadinstitute.org)



Extended Data Figure 1 | Ancestry and association summaries.

a, Multidimensional scaling plot of ancestry in the full Swedish GWAS sample, in which each point represents one individual; the left panel superimposes HapMap CEU and Finnish samples and the right panel highlights (in green) the subset of the full Swedish sample for whom we have exome sequence data. **b**, Quantile–quantile plot for gene-based SKAT results (MAF < 5% coding variants). Similar, or more conservative, profiles were obtained for other subsets of variants. **c**, Case enrichment of rare (MAF < 0.1%) and singleton disruptive mutations for the constituent sets of the primary/schizophrenia gene

set (top panel in green) and the secondary (autism/intellectual disability) gene set (bottom panel in orange). The primary set is enriched in cases (MAF < 0.1%; disruptive mutations, $P = 10^{-4}$; singletons, $P = 8 \times 10^{-4}$; significant after correction for multiple testing) whereas the autism/intellectual disability set shows only a modest trend ($P = 0.04$ for MAF < 0.1% and $P = 0.03$ for singletons) and is not significant after correction. x axis represents $-\log_{10}(P)$; OR, odds ratio. Number of genes is for total in the set (whether or not they had a rare variant).



b

Characteristic of composite set carriers (cases only)	# admissions		>0 admissions		>1 admissions		>5 admissions	
	OR	P	OR	P	OR	P	OR	P
Hospital admissions								
Duration of hospitalization for SCZ	1.02	0.610	1.02	0.512	1.02	0.625	1.02	0.663
Total number of admissions	1.00	0.957	1.00	0.968	1.02	0.789	1.02	0.690
Year of first admission	1.00	0.861	1.00	0.884	1.00	0.895	1.00	0.924
Year of most recent admission	1.01	0.230	1.01	0.246	1.01	0.235	1.01	0.254
Drug abuse	1.00	0.401	0.96	0.675	0.90	0.400	0.77	0.122
General medical condition plausibly etiologically related to psychosis	0.99	0.596	1.12	0.421	0.96	0.831	0.85	0.664
Epilepsy	1.03	0.494	0.76	0.348	0.81	0.582	1.01	0.990
Intellectual disability	1.05	0.009	1.41	0.044	1.59	0.019	2.03	0.018
Demographics								
Male	0.89	0.187	0.90	0.199	0.90	0.222	0.90	0.230
In homogeneous subset	1.11	0.287	1.10	0.323	1.10	0.316	1.11	0.280
Finnish ancestry	0.95	0.736	0.94	0.684	0.95	0.719	0.96	0.783

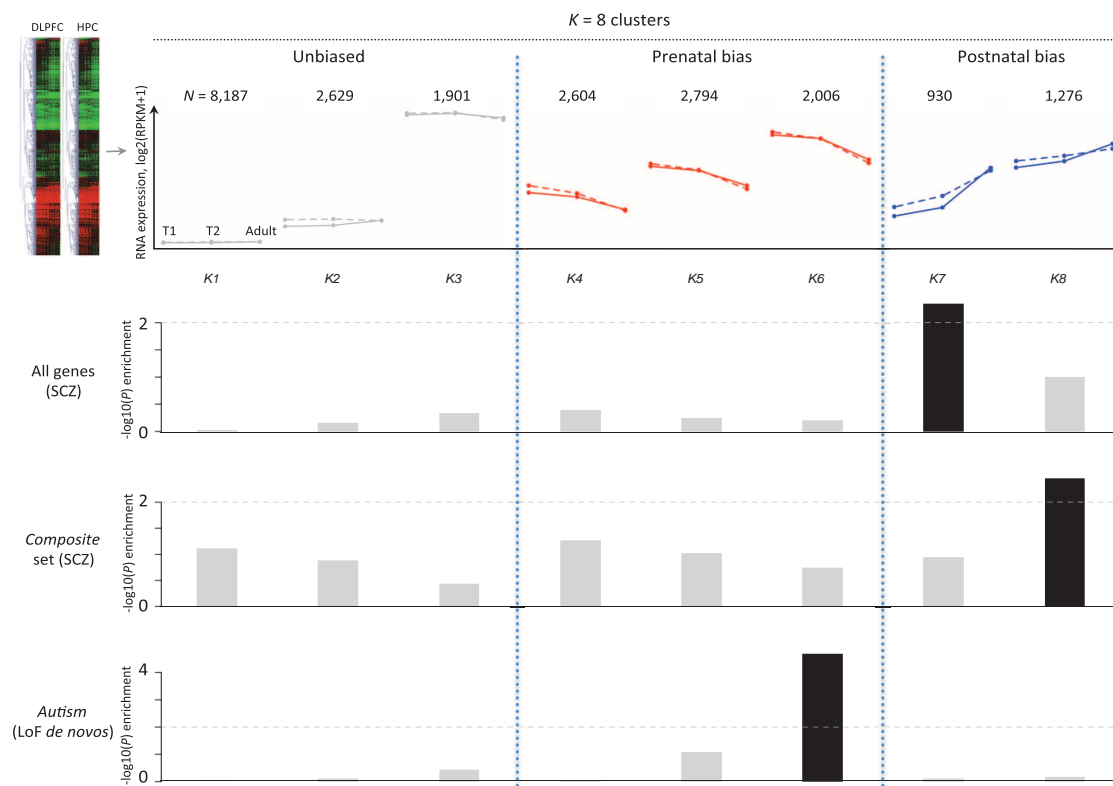
Extended Data Figure 2 | Genic and phenotypic subset analyses for the composite set. **a**, Individual gene-ranking of composite set genes. Genes are ranked by their case burden of rare disruptive mutations, from left to right, for the composite set. The squares along the bottom indicate to which sets each gene belongs. The red and blue triangles represent case and control counts for each gene. The lines above represent the statistical significance of the best test for this set: that is, the significance of the top K genes, evaluated by permutation. The black line represents results for the real data (disruptive $MAF < 0.1\%$ composite set analysis). The orange line represents the dummy condition, in which we artificially constructed a set in which the number of genes, statistical enrichment, odds ratio and case/control counts were similar to the real composite set. However, this set included the 25 top-ranked genes from individual gene-based tests (disruptive $MAF < 0.1\%$ variants), with the remainder selected at random. The profile of the best test line is markedly different between the real and dummy gene sets (note: truncated at $P = 0.0001$ reflecting the number of permutations performed). Whereas the dummy P value climbs quickly and then drops to the final aggregate result, the true composite set line continues to climb after 200 genes, indicating that many genes with a single disruptive mutation contribute to the observed set

enrichment (rather than a relatively small proportion of the 1,796 genes accounting for the majority of the signal, as in the dummy set). **b**, Phenotypic characteristics of cases carrying mutations. Relationship between clinical and demographic measures in schizophrenia cases in relation to carrying one or more composite set disruptive risk alleles ($MAF < 0.1\%$). Hospital Discharge Registry (HDR) data (ICD9 codes) were available on 979 of the 990 case carriers. All P values (uncorrected) are two-sided from a case-only joint logistic regression of carrier status (one or more risk alleles) on all admission and demographic variables including year of first and last admissions. The four pairs of columns represent analyses in which we varied the way in which the HDR admission data were represented (for drug abuse, general medication condition, epilepsy and intellectual disability). # admissions, independent variables are the untransformed number of admissions; $>X$ admissions, independent variable is binary 0/1 variable representing whether individuals had more than X admissions. Of all clinical/demographic measures considered, we observed a nominally significant increased likelihood that cases carrying a disruptive allele in the composite set have increased rates of secondary diagnoses of intellectual disability compared to other cases (based on HDR ICD9 codes).

a

		Bias: (higher in)			Prenatal			Postnatal		
		Average level:	None K1	Low/average K2	High K3	Low K4	Higher K5	Highest K6	Low K7	High K8
BrainSpan : 8-class	Geneset									
	<i>n</i> genes		8,187	2,629	1,901	2,604	2,794	2,006	930	1,276
	All genes		0.8653	0.7511	0.8719	0.4672	0.7913	0.3518	0.1639	0.0005
	Composite		0.1295	0.0447	0.3896	0.0057	0.1395	0.4188	0.1862	0.0014
	PSD genes		0.9027	0.2103	0.1043	0.7012	0.4265	0.8340	0.0244	0.0210
	FMRP targets		0.1245	0.2581	0.4230	0.0351	0.5382	0.1616	0.3797	0.0008
	SCZ <i>de novos</i> (NS)		0.1138	0.1593	0.5044	0.0089	0.1855	0.2325	0.2541	0.1367
BrainSpan : 4-class			None/low		High		Prenatal		Postnatal	
	<i>n</i> genes		2,748		3,688		3,540		1,058	
	All genes		0.8504		0.0020		0.6946		0.0255	
	Composite set		0.0322		0.0017		0.1535		0.0006	
	PSD		0.3982		0.1043		0.9692		0.0008	
	FMRP targets		0.0257		0.0100		0.1918		0.0131	
	SCZ <i>de novo</i> (NS)		0.0500		0.0145		0.0359		0.0504	
Human Brain Transcriptome :										
Xu et al. (ref. 19) 3-way classifications, HPC				Unbiased		Prenatal		Postnatal		
	<i>n</i> genes			6,578		6,841		8289		
	All genes			0.6467		0.7281		0.0517		
	Composite set			0.1009		0.2876		0.0001		
	PSD			0.3693		0.9691		0.1043		
	FMRP targets			0.5629		0.1264		0.0022		
	SCZ <i>de novo</i> (NS)			0.0397		0.1395		0.0108		
Human Brain Transcriptome :										
Xu et al. (ref. 19) 3-way classifications, DLPFC				Unbiased		Prenatal		Postnatal		
	<i>n</i> genes			6,255		6,669		8,784		
	All genes			0.5076		0.7538		0.0894		
	Composite set			0.2143		0.2054		0.0001		
	PSD			0.5949		0.8329		0.1043		
	FMRP targets			0.3205		0.4257		0.0005		
	SCZ <i>de novo</i> (NS)			0.1383		0.1189		0.0043		

b



Extended Data Figure 3 | Stratified enrichment analysis P values by developmental trajectory of expression in brain (BrainSpan and Human Brain Transcriptome (HBT) data sets). **a**, Uncorrected P values for a set of exploratory analyses in which we stratified genes in the enrichment analyses by their developmental profile of brain expression. We used four schemes to classify genes as 'brain expressed' and/or 'biased' with respect to prenatal or postnatal expression (see Supplementary Information section 6 for details). We merged data on the hippocampus and dorsolateral prefrontal cortex for the BrainSpan classifications; to mirror the classification of Xu *et al.*¹⁹ we kept separate these two groupings for the HBT data set. Results presented for MAF < 0.1% disruptive variants; similar results were obtained for singletons with the exception that the 'K4' prenatal enrichment signals were no longer significant. In general, the most consistent enrichment across variant classes, classification schemes and brain regions emerges for postnatally biased genes with high brain expression. **b**, Analysis of exome variants by developmental expression trajectory in human brain. Genes are grouped by

cluster analysis of human postmortem brain expression into eight developmental trajectories, using RNA-sequencing data from the BrainSpan project. The top row gives the number of genes per cluster and the cluster centres in log₂-scaled RPKM (reads per kilobase per million) values; solid and dotted solid lines indicate dorsolateral prefrontal cortex (DLPFC) and hippocampus (HPC), respectively. The bottom two rows show enrichment in the current study, relative to the exome-wide average, for singleton disruptive mutations in cases compared to controls, either subsetting all genes by expression profile (first row), or considering only genes in the composite set (second row). In both cases, we only observed nominally ($P < 0.01$) significant enrichment for genes that are postnatally biased. By contrast, a list of genes with loss-of-function (LoF) *de novo* mutations (compiled and reported in Fromer *et al.*³⁰) shows strong enrichment for prenatal bias (see Fromer *et al.*³⁰ for details on how *de novo* enrichment was calculated). Alternative approaches to classifying genes as prenatally or postnatally biased led to similar conclusions (Supplementary Information section 6).

Extended Data Table 1 | Sample and detected variant properties

a

Case/control status by sex ($P = 5e-11$)	Status	Male	Female	Female (%)
	Control	1291	1252	49%
	Case	1520	1016	40%
Case/control status by ancestry ($P = 2e-12$)	Ancestry	Control	Case	Case (%)
	Swedish	2356	2197	48%
	Finnish	139	274	66%

b

Sample and sequencing metrics	Cases	Controls	P (case vs. control)
n	2536	2543	.
n (pre-QC)	2546	2545	.
Total number of reads	100,532,755	100,079,333	0.62
Filtered, unique reads aligned	68,940,753	68,339,964	0.26
Filtered, unique bases aligned	5,106,614,996	5,070,497,844	0.34
Mean target coverage	89.98	89.55	0.53
Percentage of target bases covered > 10x	92.83	92.85	0.55
Percentage of target bases covered > 20x	87.30	87.30	0.93
Percentage of target bases covered > 30x	81.13	81.07	0.63
Percentage of targets w/out any bases covered at 2x	1.72	1.72	0.60
Mean number of non-reference genotypes per individual (unfiltered)	18772.9	18786.6	0.13
Mean number of on-target singletons per individual (unfiltered)	49.6	49.0	0.38
Mean dbSNP % per individual	98.3970%	98.3969%	1.00

c

Property	Variant type	n	Mean MAC	% singleton
All alternate alleles		635,944	103.37	56%
Functional class				
Noncoding		61,416	142.03	53%
Silent		185,336	152.85	51%
Missense		342,561	69.52	58%
Non-essential splice site		25,450	127.04	54%
Nonsense		9,022	20.68	69%
Essential splice-site		4,394	16.18	70%
Frameshifting indel		3,461	9.46	79%
<i>In silico</i> annotation of missenses				
LRT		168,437	34.55	62%
MutationTaster		167,316	19.90	63%
PolyPhen2 (HumDiv)		130,719	28.84	62%
PolyPhen2 (HumVar)		91,156	24.74	64%
SIFT		140,345	43.85	61%
Primary variant groupings for analysis				
Singletons	Gene disruptive	12,047	1.00	100%
	NSstrict	36,542	1.00	100%
	NSbroad	160,229	1.00	100%
<0.1% MAF (1-10 alleles)	Gene disruptive	15,972	1.56	75.4%
	NSstrict	50,369	1.65	72.5%
	NSbroad	233,575	1.78	68.6%
<0.5% MAF (1-50 alleles)	Gene disruptive	16,523	2.24	72.9%
	NSstrict	52,545	2.51	69.5%
	NSbroad	248,217	3.04	64.6%

a. Numbers of individuals in the final data set, after individual-level QC. Finnish ancestry was inferred by multidimensional scaling. P values from Fisher's exact test. **b.** Technical metrics for the cases and controls (after individual-level QC); P values for two-sided test of case/control differences (t -test). **c.** Properties of variants detected by exome sequencing. Counts (N) and minor allele counts (MAC) for various classes of variant in the main exome data set, following all QC. Missense deleteriousness prediction algorithms and how they were combined are described in Supplementary Information section 4.

Extended Data Table 2 | Genes prioritized as more likely to harbour large-effect alleles

Class	Gene	Singletons	MAF < 0.1%	Notes
ARC/PSD complex				
	<i>CYFIP1</i>	1/0	1/0	SCZ <i>de novo</i> (CNV)
	<i>BAIAP2</i>	1/0	1/0	SCZ <i>de novo</i> (NS)
	<i>DLG1</i>	1/0	1/0	SCZ <i>de novo</i> (NS), SCZ <i>de novo</i> (CNV)
	<i>SLC25A3</i>	1/0	1/0	
	<i>GLUD1</i>	1/0	1/0	
	<i>CAMK2A</i>	1/0	1/0	FMRP target
	<i>ATP1B1</i>	1/0	1/0	AUT <i>de novo</i> (disruptive); FMRP target
	<i>IQSEC2</i>	1/0	1/0	ID <i>de novo</i> (disruptive); FMRP target
	<i>MBP</i>	1/0	1/0	FMRP target
	<i>Total</i>	9/0	9/0	
		$P = 0.0016$	$P = 0.0014$	
		OR = 19.2	OR = 19.2	
		(2.4 - 2471)	(2.4 - 2471)	
PSD-95 geneset				
	<i>ABLIM1</i>	1/0	1/0	
	<i>ACO2</i>	1/0	1/0	FMRP target
	<i>ANKS1B</i>	3/1	3/1	
	<i>ATP1B1</i>	1/0	1/0	AUT <i>de novo</i> (disruptive); FMRP target
	<i>ATP5A1</i>	1/0	1/0	FMRP target
	<i>BAIAP2</i>	1/0	1/0	SCZ <i>de novo</i> (NS)
	<i>CAMK2A</i>	1/0	1/0	FMRP target
	<i>CAMK2B</i>	2/0	2/0	FMRP target
	<i>DLG1</i>	1/0	1/0	SCZ <i>de novo</i> (NS), SCZ <i>de novo</i> (CNV)
	<i>GAPDH</i>	1/0	1/0	
	<i>IQSEC2</i>	1/0	1/0	ID <i>de novo</i> (disruptive); FMRP target
	<i>NRXN1</i>	1/0	1/0	SCZ <i>de novo</i> (NS); AUT <i>de novo</i> (disruptive); FMRP target
	<i>PRDX1</i>	0/1	0/1	
	<i>PRDX2</i>	0/1	0/1	
	<i>SUCLA2</i>	1/0	1/0	AUT <i>de novo</i> (disruptive)
	<i>SYNGAP1</i>	1/0	1/0	SCZ <i>de novo</i> (disruptive); ID <i>de novo</i> (disruptive); FMRP target
	<i>Total</i>	17/3	17/3	
		$P = 0.0017$	$P = 0.0009$	
		OR = 5.1 (1.8 - 19.2)	OR = 5.1 (1.8 - 19.2)	
Voltage-gated calcium ion channel genes				
	<i>CACNA1B</i>	1/0	1/0	FMRP target
	<i>CACNA1C</i>	2/0	2/0	SCZ & BP GWAS hit
	<i>CACNA1H</i>	1/0	3/0	
	<i>CACNA1S</i>	2/0	2/3	SCZ & AUT <i>de novos</i> (NS)
	<i>CACNA2D1</i>	1/0	1/0	PSD
	<i>CACNA2D2</i>	3/0	3/0	
	<i>CACNA2D3</i>	0/0	3/0	AUT <i>de novo</i> (disruptive)
	<i>CACNA2D4</i>	1/0	2/4	
	<i>CACNB2</i>	0/1	0/1	
	<i>CACNB4</i>	1/0	1/0	PSD
	<i>Total</i>	12/1	15/8	
		$P = 0.0021$	$P = 0.021$	
		OR = 8.4 (2.03 - 77)	OR = 2.1 (0.97 - 4.9)	
Top disruptive gene-based test				
	<i>KYNU</i>	3/0	10/0	
		$P = 0.13$	$P = 0.0017$	
			OR = 21.2 (2.7 - 2725)	

Individual gene case/control counts, odds ratios and *P* values for genes from primary gene sets with odds ratios >5, and *KYNU* (top-ranked individual gene). Odds ratios are calculated using Firth's method (penalized maximum likelihood logistic regression) and shown with 95% confidence intervals. *P* values are empirical, uncorrected one-sided burden tests. FMRP target annotations are based on the Darnell *et al.*⁴⁰ list only. Supplementary Table 1 lists singleton variant and genotype information for the genes listed here.

Extended Data Table 3 | Extended results for all PSD gene sets

a

Set	<i>n</i> genes	Disruptive			NS _{strict}			NS _{broad}		
		Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%
PSD (human core)	685	0.0729	0.1019	0.1083	0.0058	0.1045	0.1285	0.0866	0.5827	0.3743
ARC	28	0.0016	0.0013	0.0014	0.0004	0.0018	0.0047	0.2830	0.4607	0.3542
NMDAR network	61	0.0154	0.0229	0.0225	0.0001	0.0007	0.0005	0.0012	0.1426	0.0420
mGluR5	39	0.1299	0.0861	0.0862	0.0628	0.0837	0.0900	0.0302	0.2192	0.1329
PSD-95	65	0.0015	0.0008	0.0008	0.0027	0.0204	0.0393	0.2992	0.3722	0.1147
Pre-synapse	431	0.0187	0.0983	0.1458	0.2327	0.1811	0.3600	0.1306	0.7597	0.6515
Pre-synaptic active zone	173	0.0518	0.0487	0.0482	0.6162	0.6641	0.7082	0.8439	0.9918	0.9554
Synaptic vesicle	344	0.1030	0.3133	0.4151	0.1439	0.1093	0.2466	0.0375	0.3423	0.2718
Cytoplasm	271	0.5851	0.1793	0.1034	0.8983	0.5351	0.6007	0.1861	0.1084	0.1192
Early Endosomes	17	0.8917	0.7860	0.7826	0.2891	0.2420	0.2139	0.1472	0.4019	0.4123
Endoplasmic Reticulum	97	0.3005	0.1882	0.2612	0.6615	0.2805	0.5036	0.6194	0.5044	0.7878
ER/Golgi-derived vesicles	94	0.4258	0.2678	0.3644	0.3001	0.4977	0.6239	0.3063	0.6944	0.7649
Golgi	31	0.5130	0.5493	0.5481	0.1998	0.0921	0.1338	0.0074	0.1628	0.4178
Mitochondrion	197	0.0141	0.0259	0.0178	0.4351	0.0860	0.0671	0.6999	0.2112	0.6079
Nucleus	167	0.1790	0.3029	0.2900	0.0626	0.1512	0.2728	0.2006	0.3340	0.3196
Plasma membrane	50	0.7940	0.5659	0.5635	0.9416	0.8059	0.8091	0.8944	0.5531	0.3028
Recycling Endosomes/trans-Golgi network	68	0.1502	0.0944	0.1556	0.5349	0.4514	0.5359	0.0902	0.0862	0.1862

b

	PSD (human core) (ref. 13)			FMRP targets (Darnell, ref. 40)		FMRP targets (Ascano, ref. 42)	
	<i>n</i>	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
PSD (human core)	685	.	.	170	25%	80	12%
ARC	28	.	.	16	57%	2	7%
NMDAR network	61	.	.	32	52%	5	8%
mGluR5	39	.	.	25	64%	7	18%
PSD-95	65	.	.	30	46%	4	6%
Pre-synapse	431	213	49%	87	20%	31	7%
Pre-synaptic active zone	173	121	70%	50	29%	10	6%
Synaptic vesicle	344	162	47%	72	21%	27	8%
Cytoplasm	271	77	28%	16	6%	23	8%
Early Endosomes	17	6	35%	2	12%	1	6%
Endoplasmic Reticulum	97	13	13%	5	5%	4	4%
ER/Golgi-derived vesicles	94	24	26%	7	7%	4	4%
Golgi	31	2	6%	2	6%	4	13%
Mitochondrion	197	57	29%	6	3%	12	6%
Nucleus	167	19	11%	7	4%	19	11%
Plasma membrane	50	16	32%	6	12%	5	10%
Recycling Endosomes/trans-Golgi network	68	19	28%	3	4%	7	10%
Total	1509	685	45%	170	11%	80	5%

a, Full PSD gene set association results. For all nine (three annotation levels by three frequency levels), *P* values for enrichment of all gene sets described and tested in Kirov *et al.*¹³. In addition to the PSD genes (top five rows), enrichment statistics for presynaptic genes, and neuronal genes clustered on the basis of subcellular location are given. Although the *P* values presented are uncorrected, we performed this analysis correcting for all $9 \times 17 = 153$ tests (by considering the distribution of the minimum empirical *P* value across tests and sets, as described in the Supplementary Information). The values in bold are significant ($P_{\text{corrected}} < 0.05$) after correction for multiple testing. Both ARC and NMDAR network are significant after multiple test correction, for the singleton NS_{strict} category. (Note: for ARC the disruptive singleton category is, as reported in the primary test, highly significant and withstands correction for multiple testing in that context; in this broader, less focused analysis it yields $P_{\text{corrected}} = 0.17$; the majority of $P_{\text{corrected}}$ values (not shown) are 1.00.) **b**, PSD and FMRP-target gene sets: descriptive statistics and overlap. Overlap between Darnell *et al.*⁴⁰ and Ascano *et al.*⁴² FMRP targets and PSD genes: for example, 57% (16 out of 28) of ARC genes are in the Darnell FMRP list. By contrast, only 7% (2 out of 28) are in the Ascano list. There is a similar trend across the three other major PSD subsets considered here: NMDAR network, PSD-95 and mGluR5 genes. Conversely, 22% of Darnell targets are in the PSD (human core) compared to only 9% of Ascano targets.

Extended Data Table 4 | Association results for individual CNV regions

Group	Genes	Disruptive		NS _{strict}	
		Singletons	MAF < 0.1%	Singletons	MAF < 0.1%
CNV loci	All	0.3279	0.4557	0.0843	0.0044
	1q21.1	0.4533	0.6966	0.3205	0.1832
	2p16.3	0.4775	0.3580	0.4703	0.2750
	3q29	0.1054	0.0068	0.0123	0.0006
	7q36.3	0.8642	0.5750	0.6411	0.2688
	7q11.23	0.7199	0.6800	0.3329	0.2207
	15q11.2	0.3208	0.1616	0.5138	0.1362
	15q13.3	0.0883	0.3976	0.3672	0.2746
	16p13.11	0.4194	0.3775	0.8346	0.4124
	16p11.2	0.1613	0.1240	0.0655	0.0974
	17q12	0.7377	0.5205	0.4313	0.1385
	22q11.21	0.9386	0.9977	0.5456	0.8754
Gene		A/U	Gene name		
3q29 genes	<i>DLG1</i>	5/0	discs, large homolog 1 (Drosophila)		
	<i>RNF168</i>	5/1	ring finger protein 168, E3 ubiquitin protein ligase		
	<i>CEP19</i>	2/0	centrosomal protein 19kDa		
	<i>LRRC33</i>	2/0	leucine rich repeat containing 33		
	<i>PAK2</i>	2/0	p21 protein (Cdc42/Rac)-activated kinase 2		
	<i>PCYT1A</i>	5/2	phosphate cytidylyltransferase 1, choline, alpha		
	<i>PIGX</i>	6/3	phosphatidylinositol glycan anchor biosynthesis, class X		
	<i>FBXO45</i>	1/0	F-box protein 45		
	<i>NCBP2</i>	1/0	nuclear cap binding protein subunit 2, 20kDa		
	<i>PIGZ</i>	1/0	phosphatidylinositol glycan anchor biosynthesis, class Z		
	<i>TFRC</i>	1/0	transferrin receptor (p90, CD71)		
	<i>ZDHHC19</i>	1/0	zinc finger, DHHC-type containing 19		
	<i>C3orf43</i>	4/2	chromosome 3 open reading frame 43		
	<i>MFI2</i>	15/12	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5		
	<i>SLC51A</i>	1/1	solute carrier family 51, alpha subunit		
	<i>BDH1</i>	0/1	3-hydroxybutyrate dehydrogenase, type 1		
	<i>TCTEX1D2</i>	0/1	Tctex1 domain containing 2		
	<i>WDR53</i>	0/1	WD repeat domain 53		

Focused enrichment analysis of genes under schizophrenia-associated CNV regions. The top panel presents omnibus P values testing all genes/regions (bold indicates significance after correction for the four tests, $p_{\text{corrected}} = 0.016$ for NS_{strict} MAF < 0.1% variants). This enrichment arises solely from the 3q29 locus (middle panel; bold indicates significance after correction of the 44 tests performed, $p_{\text{corrected}} = 0.024$ for 3q29). Genes and NS_{strict} case/control counts for the 3q29 region (bottom panel).

Extended Data Table 5 | Further stratification of enrichment analyses by class of variant

a

Primary geneset	P	Disruptive singletons			OR	Damaging missense (strict) singleton			
		n	A/U			P	n	A/U	OR
All primary genes	0.0008	905	852/716	1.20		0.0393	1357	2080/2001	1.04
SCZ <i>de novo</i> genes (refs. 18,19,30)									
Exome sequencing (disruptive)	0.0349	40	56/38	1.48		0.5613	53	121/120	1.01
Exome sequencing (nonsyn)	0.0059	332	384/309	1.25		0.2776	393	750/736	1.02
Copy number variants (refs. 5,13)									
<i>de novo</i> CNV genes	0.0224	64	61/40	1.53		0.0593	90	125/112	1.12
SCZ-associated CNV genes	0.3378	72	65/55	1.19		0.0310	111	148/119	1.25
GWAS (refs. 3,5,12)									
Voltage-gated calcium channel genes	0.0021	9	12/1	8.40		0.4629	18	37/35	1.06
Common SNPs (P < 1e-4 intervals)	0.1832	185	165/146	1.14		0.9246	268	359/395	0.91
miR-137 targets	0.6643	140	98/100	0.99		0.1415	263	376/361	1.05
Synaptic genes (ref. 13)									
PSD (human core)	0.0824	219	172/145	1.19		0.0070	394	646/581	1.12
ARC	0.0012	9	9/0	19.20		0.0069	19	32/15	2.14
NMDAR network	0.0162	17	17/5	3.42		0.0003	34	76/45	1.70
PSD-95	0.0018	16	17/3	5.10		0.0218	34	44/30	1.47
mGluR5	0.1335	10	9/3	3.02		0.1715	22	52/36	1.45

b

Geneset	Case-unique burden analysis			Known variants				Novel variants			
	P	n genes(A/U)	A(U)	Singletons		MAF < 0.1%		Singletons		MAF < 0.1%	
				P	n	P	n	P	n	P	n
Composite	0.0006	829(275/214)	378(297)	0.3733	145	0.2202	226	0.0002	683	0.0005	744
Primary	0.0022	1026(325/265)	440(367)	0.1003	191	0.0417	299	0.0074	831	0.0058	910
SCZ <i>de novo</i> genes (refs. 18,19,30)											
Exome sequencing (disruptive)	0.0018	47(16/6)	29(12)	0.5514	13	0.2196	24	0.0362	35	0.0010	40
Exome sequencing (nonsyn)	0.0037	371(108/80)	159(116)	0.3647	94	0.2064	144	0.0142	302	0.0071	326
Copy number variants (refs. 5,13)											
<i>de novo</i> CNV genes	0.1267	79(25/17)	32(24)	0.0156	13	0.0116	24	0.0819	59	0.0679	67
SCZ-associated CNV genes	0.7971	90(20/23)	24(32)	0.0355	23	0.0069	34	0.6081	63	0.9187	76
GWAS (refs 3,5,12)											
Voltage-gated calcium channel	0.0129	9(5/1)	10(1)	0.4922	2	0.7077	3	0.0006	7	0.0022	8
P < 1e-4 intervals	0.1079	211(65/55)	91(78)	0.0394	44	0.0409	69	0.4425	164	0.1882	180
miR-137 targets	0.4498	156(52/50)	67(60)	0.9939	14	0.9972	22	0.3846	133	0.2757	147
Synaptic genes (ref. 13)											
PSD (human core)	0.2234	244(92/79)	113(109)	0.5348	25	0.3072	40	0.1091	205	0.1629	226
ARC	0.0008	9(9/0)	9(0)	.	0	.	0	0.0016	9	0.0013	9
NMDAR network	0.0105	21(13/4)	18(4)	1.0000	1	0.6905	2	0.0075	16	0.0085	19
PSD-95	0.0137	16(13/2)	14(2)	0.1218	1	0.1559	1	0.0034	15	0.0022	15
mGluR5	0.1363	11(7/1)	8(1)	.	0	0.1458	1	0.1427	10	0.1826	11
Secondary (autism/ID)	0.0916	1249(348/314)	479(471)	0.1679	226	0.3543	352	0.1834	1041	0.0807	1143
<i>De novo</i> genes (exome sequencing)											
Autism (disruptive) (refs. 22-25)	0.662	65(17/17)	20(26)	0.4463	18	0.0781	23	0.6161	50	0.7009	56
Autism (nonsyn) (refs. 22-25)	0.220	407(101/96)	143(154)	0.3198	89	0.4487	133	0.6656	336	0.4960	369
ID (disruptive) (refs. 20, 21)	0.262	8(4/1)	4(2)	1.0000	1	0.2747	3	0.3558	8	0.0578	8
ID (nonsyn) (refs. 20,21)	0.052	69(22/18)	35(28)	0.1303	14	0.0934	26	0.5331	62	0.3368	66
Neurodevelopmental candidates											
ASD candidates (ref. 39)	0.110	37(12/6)	16(7)	0.5824	9	0.7543	14	0.0484	24	0.0429	29
ID candidates (ref. 39)	0.994	88(14/28)	16(38)	0.6553	16	0.7056	24	0.9488	74	0.9556	82
Autism PPI networks											
CHD8 network (ref. 24)	1.000	1(0/1)	0(1)	.	0	.	0	1.0000	1	1.0000	1
49-gene network (ref. 24)	0.796	19(3/7)	4(16)	0.4755	5	0.7326	7	0.6081	30	0.7231	33
74-gene network (ref. 24)	0.654	33(6/13)	10(28)	0.6438	4	0.6667	4	0.7285	17	0.8411	19
Fragile X mental retardation protein targets											
Darnell (ref. 40) targets	0.022	341(131/95)	169(133)	0.3048	39	0.3889	61	0.0007	288	0.0022	309
Ascano (ref. 42) targets	0.449	517(134/131)	187(200)	0.5571	83	0.7281	128	0.5261	439	0.4089	482
Ascano (ref. 42) FMRP/autism	0.423	33(10/6)	12(12)	0.0384	5	0.4624	11	0.6088	23	0.3954	28

a. Gene set analyses for damaging missense mutations only. For the primary gene set and the 12 constituent subsets, a comparison of disruptive versus (strictly defined) damaging missenses, that is, an independent set of variants. The omnibus result for the primary test is modest ($P = 0.04$) and did not withstand correction for multiple testing: as illustrated in Fig. 1 and the main text, the bulk of the enrichment signal we observe comes from (singleton) disruptive mutations. Nonetheless, specific gene sets such as ARC and the NMDAR network are highly and independently enriched for missense variants. *N* represents the number of genes with at least one mutation of this class observed in the sample. *A/U* represent case/control counts of non-reference genotypes. *OR* represents the odds ratio (not corrected for exome-wide rates) estimated by Firth's method for sets with small cell counts. All tests are empirical and one-sided (higher values expected in cases) as described in the main text and Methods. **b.** Enrichment analyses of novel and case-unique disruptive mutations. For primary and secondary gene sets (and constituent subsets) as well as the composite set: results of alternative burden analyses. First, focusing only on genes without any control disruptive variants; no further frequency filter is imposed. Here 'N genes(A/U)' indicates the number of genes with at least one disruptive variant, followed by the number of genes with case-only disruptive mutations and (for comparison) the number with control-only disruptive mutations. The 'A(U)' column gives the number of case variants in the case-only genes: the test statistic is based on the empirical distribution of this count. The *U* in this field represents the similar quantity for controls (not explicitly used in the statistic). The second set of analyses represent standard burden/enrichment tests (that is, as Tables 1 and 2) but stratified for novel versus known disruptive variants, according to dbSNP and the Exome Sequencing Project/Exome Variant Server (ESP/EVS) database. Novel variants show greater enrichment, although most rare variants observed in our study (both in cases and in controls) are novel, so tests of novel variants will have greater power.

Extended Data Table 6 | Gene set analysis of *de novo* genes from schizophrenia exome-sequencing studies

a

Set	<i>n</i> genes	Disruptive			NS _{strict}			NS _{broad}		
		Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%
Fromer et al. (ref. 30) disruptive	63	0.1484	0.0075	0.0034	0.7401	0.7324	0.6264	0.3347	0.1660	0.2536
Fromer et al. (ref. 30) nonsyn	464	0.0004	0.0003	0.0016	0.0341	0.0057	0.0892	0.4688	0.4842	0.6547
Girard & Xu (refs. 18,19) disruptive	24	0.0342	0.0082	0.0082	0.0423	0.0412	0.0602	0.0774	0.1106	0.0891
Girard & Xu (refs. 18,19) nonsyn	151	0.6916	0.4124	0.4186	0.1510	0.1326	0.1285	0.2258	0.3420	0.1790
Combined SCZ disruptive	87	0.0319	0.0007	0.0003	0.3355	0.3162	0.2692	0.1325	0.0757	0.1016
Combined SCZ nonsyn	611	0.0053	0.0011	0.0055	0.0192	0.0024	0.0379	0.3408	0.4064	0.4472

b

Gene	<i>De novo</i> study (type)	Test	<i>n</i>	A/U	<i>P</i>	Gene name
<i>ALDH1L2</i>	Fromer (ref. 30) nonsyn	disruptive	6	10/3	0.028	aldehyde dehydrogenase 1 family, member L2
<i>CACNA1S</i>	Fromer (ref. 30) nonsyn	NS _{strict}	23	28/15	0.031	calcium channel, voltage-dependent, L type, alpha 1S subunit
<i>DLG1</i>	Fromer (ref. 30) nonsyn	NS _{strict}	4	5/0	0.021	discs, large homolog 1 (Drosophila)
<i>IGSF22</i>	Fromer (ref. 30) nonsyn	disruptive	5	5/0	0.043	immunoglobulin superfamily, member 22
<i>JARID2</i>	Fromer (ref. 30) nonsyn	NS _{strict}	5	5/0	0.041	jumonji, AT rich interactive domain 2
<i>LAMA4</i>	Fromer (ref. 30) nonsyn	NS _{strict}	8	12/4	0.041	laminin, alpha 4
<i>NBEA</i>	Fromer (ref. 30) nonsyn	NS _{strict}	5	5/0	0.025	neurobeachin
<i>POLL</i>	Fromer (ref. 30) nonsyn	disruptive	4	4/0	0.042	polymerase (DNA directed), lambda
<i>PTK2B</i>	Fromer (ref. 30) nonsyn	NS _{strict}	4	4/0	0.044	PTK2B protein tyrosine kinase 2 beta
<i>SHKBP1</i>	Fromer (ref. 30) nonsyn	NS _{strict}	9	15/4	0.018	SH3KBP1 binding protein 1
<i>SULF2</i>	Fromer (ref. 30) nonsyn	NS _{strict}	5	7/0	0.007	sulfatase 2
<i>SYNGAP1</i>	Xu (ref. 19) LoF	NS _{strict}	4	4/0	0.043	synaptic Ras GTPase activating protein 1
<i>SZT2</i>	Xu (ref. 19) LoF	NS _{strict}	22	18/9	0.049	seizure threshold 2 homolog (mouse)
<i>TANC1</i>	Fromer (ref. 30) nonsyn	NS _{strict}	14	17/4	0.002	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
<i>TEP1</i>	Xu (ref. 19) nonsyn	NS _{strict}	26	25/14	0.048	telomerase-associated protein 1
<i>UFL1</i>	Fromer (ref. 30) LoF	NS _{strict}	4	7/0	0.008	UFM1-specific ligase 1
<i>UFL1</i>	Fromer (ref. 30) LoF	disruptive	2	5/0	0.029	UFM1-specific ligase 1

a. Test of case enrichment of rare variants in cases compared to controls, for genes with one or more *de novos* in Fromer *et al.*³⁰, Xu *et al.*¹⁹ and/or Girard *et al.*¹⁸. The *P* values in bold are significant at $p_{\text{corrected}} < 0.05$, correcting for all $3 \times 3 \times 6 = 54$ tests reported. **b.** Genes nominally significant (no correction) that had an observed *de novo* in one of the schizophrenia studies.

Extended Data Table 7 | Summary of observed likely N.A, deleterious variants in ARC genes across studies

ARC gene (<i>n</i> =28)	Current study		<i>de novo</i> CNV (Kirov et al., ref. 13)	<i>de novo</i> SNV (Fromer et al., ref. 30)	<i>de novo</i> SNV in ID (refs. 20,21)
	Disruptive	Damaging missense (strict)			
ACTN4		3/1			
ARF5					
ATP1A1		3/0			
ATP1A3		2/1			
ATP1B1	1/0	1/0			
BAIAP2	1/0			NS(x2)	
CAMK2A	1/0	1/0			
CRMP1		1/3			
CYFIP1	1/0	4/1	2 del; 2 dup		
DLG1	1/0	2/0	1 del	NS	
DLG2		2/3	2 del	LoF	
DLG4		1/2			NS
DLGAP1		2/0	1 del		
DLGAP2		1/0			
DPYSL2		0/1			
GLUD1	1/0	1/0			
GLUL		2/0			
GRIN1		2/0			
HSPA8				LoF & NS	
IQSEC1		4/1			
IQSEC2	1/0	0/1			LoF
MBP	1/0	0/1			
PKM2					
PLP1					
SLC25A3	1/0				
SLC25A4					
SLC25A5					
STXBP1					LoF, NS(x2)
Counts:	9 / 0	32 / 15	8 CNVs	6 SNVs	5 SNVs
P-value:	0.0016	0.0069	0.00025	0.0005	0.00002

For the 28 ARC genes, a summary of which genes had singleton disruptive, or damaging missense, variants in the current study, compiled alongside the genes with *de novo* CNVs or SNVs observed in Kirov *et al.*¹³ or Fromer *et al.*³⁰ as well as the intellectual disability (ID) *de novo* genes (compiled in Fromer *et al.*³⁰). The *P* values at the bottom indicate that in each comparison the ARC gene set was significantly enriched.

Molecular control of δ -opioid receptor signalling

Gustavo Fenalti^{1*}, Patrick M. Giguere^{2*}, Vsevolod Katritch¹, Xi-Ping Huang², Aaron A. Thompson¹, Vadim Cherezov¹, Bryan L. Roth² & Raymond C. Stevens¹

Opioids represent widely prescribed and abused medications, although their signal transduction mechanisms are not well understood. Here we present the 1.8 Å high-resolution crystal structure of the human δ -opioid receptor (δ -OR), revealing the presence and fundamental role of a sodium ion in mediating allosteric control of receptor functional selectivity and constitutive activity. The distinctive δ -OR sodium ion site architecture is centrally located in a polar interaction network in the seven-transmembrane bundle core, with the sodium ion stabilizing a reduced agonist affinity state, and thereby modulating signal transduction. Site-directed mutagenesis and functional studies reveal that changing the allosteric sodium site residue Asn 131 to an alanine or a valine augments constitutive β -arrestin-mediated signalling. Asp95Ala, Asn310Ala and Asn314Ala mutations transform classical δ -opioid antagonists such as naltrindole into potent β -arrestin-biased agonists. The data establish the molecular basis for allosteric sodium ion control in opioid signalling, revealing that sodium-coordinating residues act as ‘efficacy switches’ at a prototypic G-protein-coupled receptor.

The three classical opioid receptors (μ , κ and δ -OR) and the related nociceptin/orphanin FQ peptide receptor (NOP) are G-protein-coupled receptors (GPCRs) essential for regulating nociception, mood and awareness¹. These opioid GPCRs are activated by endogenous peptides (endorphins, enkephalins, dynorphins, nociceptin/orphanin FQ), natural alkaloids (opiates), and an expanding number of small molecule agonists through interactions with the orthosteric site located in the extracellular portion of the seven-transmembrane (7TM) bundle. Despite the progress made in understanding GPCR activation², the underlying molecular mechanisms and structural features responsible for many processes including signal transduction, allosteric modulation, functional selectivity and constitutive activity remain elusive^{3,4}.

Insights from 1.8 Å resolution δ -OR structure

Pioneering studies initiated in 1973 on opioid receptors revealed that physiological concentrations of sodium alter opiate ligand binding and signalling, albeit by unknown mechanisms^{5,6}. To address the molecular basis for the striking allosteric effect of sodium on opioid receptor function, we crystallized the human δ -OR (residues 36–338) with an amino-terminal *b*₅₆₂RIL (BRIL) fusion protein (BRIL- δ OR(Δ N/ Δ C)) and determined the crystal structure in complex with the subtype-selective ligand naltrindole⁷ at 1.8 Å resolution (Fig. 1, Extended Data Table 1 and Methods). Importantly, the high-resolution BRIL- δ OR(Δ N/ Δ C)-naltrindole structure contains the wild-type protein sequence, including an intact intracellular loop 3 (ICL3), providing the opportunity to study an opioid receptor that closely resembles a near native conformational state.

The 1.8 Å structure of the human δ -OR is similar to the 3.4 Å *Mus musculus* δ -OR structure fused to T4 lysozyme⁸ at the ICL3 site (root mean squared deviation of 0.91 Å over all structurally characterized C α atoms) with the distinction that the atomic details of regions crucial for receptor activity are revealed. These include: (1) a fully resolved ICL3 adopting a ‘closed’ inactive state conformation (Fig. 2); (2) a detailed molecular characterization of the orthosteric site with water-mediated ligand–receptor interactions (Extended Data Fig. 1); (3) a

distinct conformation of the human third extracellular loop (ECL3) (Extended Data Fig. 2); and, importantly, (4) a high-resolution characterization of the allosteric sodium site, water molecules and a comprehensive network of hydrogen-bond interactions inside the 7TM core (Fig. 1 and Extended Data Figs 3, 4).

All ICL3 residues are well resolved in the BRIL- δ OR(Δ N/ Δ C)-naltrindole structure. The side-chain guanidinium group of Arg 257^{6,31} (superscripts indicate residue numbering using the Ballesteros–Weinstein nomenclature⁹) appears to have a key role in stabilizing ICL3 by forming an extensive hydrogen-bonding network with the main-chain carbonyls of Leu 240^{5,67}, Arg 244^{ICL3} and Val 243^{ICL3}, and a salt bridge with the carboxylate group of Asp 253^{6,27} (Fig. 2a). The Leu 246^{ICL3} and Val 243^{ICL3} side chains insert back in the helical bundle and form a hydrophobic cluster with Val 150^{3,54}, Leu 240^{5,67} and Leu 256^{6,30} (Fig. 2b). The loop also interacts with helix III via a water-mediated hydrogen-bond network between the main-chain carbonyl groups of Leu 246^{ICL3} and Val 150^{3,54}, and the side chain of Arg 239^{5,66} (Fig. 2d). These atomic details suggest a stable ‘closed’ conformation of ICL3 in the inactive δ -OR, which tethers the intracellular ends of helices V and VI. Although it contrasts with the more ‘exposed’ ICL3 conformations in the thermally stabilized A_{2A} adenosine receptor (A_{2A}AR; Protein Data Bank (PDB) accession 3PWH)¹⁰ and rhodopsin (PDB 3CAP)¹¹ (Fig. 2c), the ICL3 in δ -OR is similar to that observed in the lower resolution NOP structure (PDB 4EA3)¹² (Fig. 2b). A high sequence conservation of ICL3 in all four opioid receptors, which signal primarily via G $\alpha_{i/o}$ -proteins, suggests that ICL3 can adopt a similar ‘closed’ conformation in inactive states of all opioid receptor subtypes. The closed conformation of ICL3 may have a role in stabilizing the inactive state in opioid receptors, and thus compensate for the lack of a stabilizing ‘ionic lock’ in these receptors, which have a hydrophobic Leu^{6,30} instead of the usual Glu^{6,30} side chain that is required for an ionic lock.

In the orthosteric pocket, the BRIL- δ OR(Δ N/ Δ C)-naltrindole structure reveals an extensive network of water-mediated interactions with the morphinan group of naltrindole, including interactions with residues

¹Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²National Institute of Mental Health Psychoactive Drug Screening Program and Department of Pharmacology and Division of Chemical Biology and Medicinal Chemistry, University of North Carolina Chapel Hill Medical School, Chapel Hill, North Carolina 27599, USA.

*These authors contributed equally to this work.

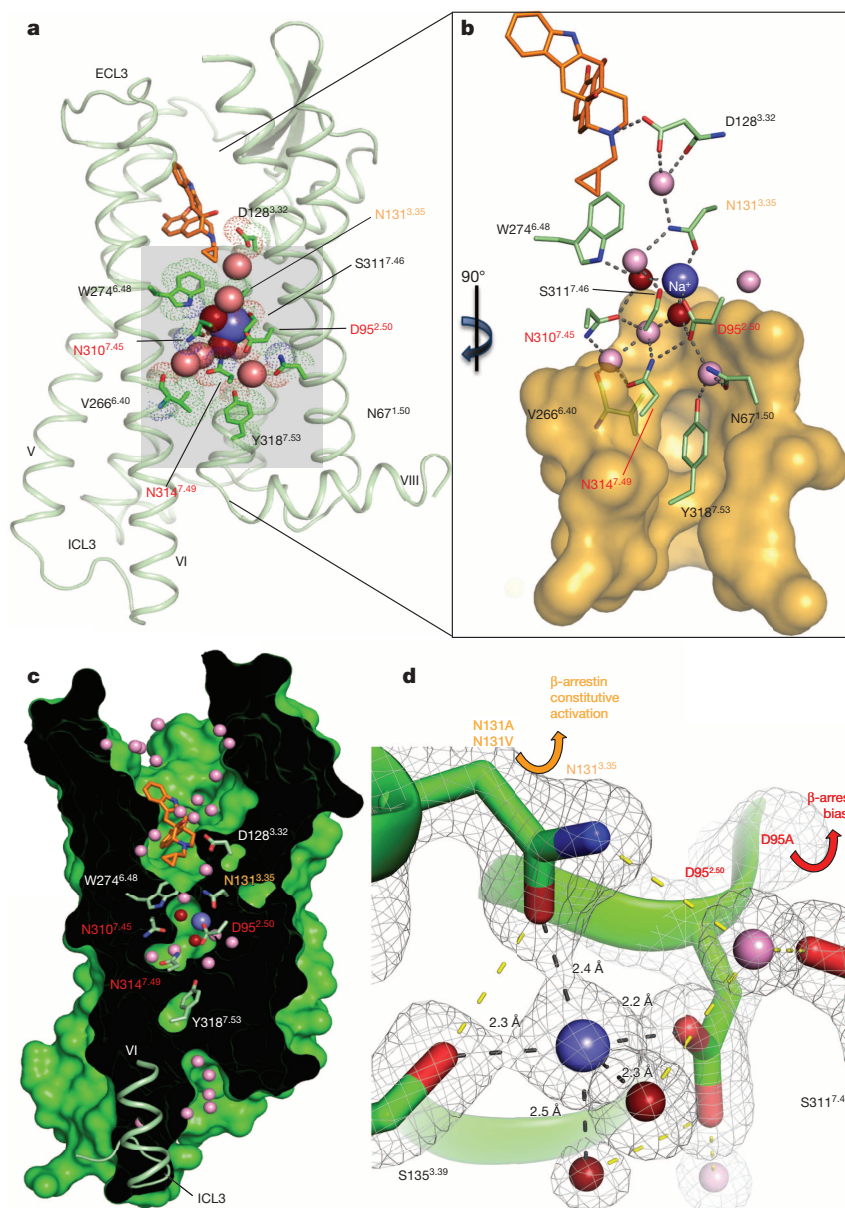


Figure 1 | Interactions in the 7TM core of BRIL- δ OR(Δ N/ Δ C)-naltrindole. **a**, BRIL- δ OR(Δ N/ Δ C)-naltrindole structure (light green, BRIL fusion omitted) and residues around the allosteric sodium site (green sticks). Sodium is shown as a blue sphere; red and pink spheres are waters in the first and second coordination shells, respectively. Naltrindole is shown as orange sticks. **b**, Hydrogen bonds (grey dotted lines) and hydrophobic residues (orange surface, located below the allosteric site). **c**, 'Sliced' surface representation of BRIL- δ OR(Δ N/ Δ C)-naltrindole showing the continuous pathway

connectivity between orthosteric and allosteric sites. **d**, $2mF_o - DF_c$ electron density map (grey mesh) contoured at 2σ around residues, waters and sodium in the allosteric site. Hydrogen bonds are shown as black (first sodium ion coordination shell) and yellow (other hydrogen bonds) dotted lines. Arrows indicate increased β -arrestin constitutive activity of 'efficacy switch' mutants Asn131^{3.35}Ala and Asn131^{3.35}Val (yellow) and β -arrestin-biased activation in response to naltrindole in the Asp95^{2.50}Ala mutant (red).

in helix V and ECL2 (Extended Data Fig. 1a). For the ECL3 region, a key selectivity determinant for peptide binding to classical opioid receptors¹³, we observe that the side chain of Arg 291^{ECL3} constrains a distinct loop conformation between helices VI and VII through hydrogen-bonding networks with the main-chain carbonyl groups of Val 287^{ECL3} and Trp 284^{6.58}, positioning the latter for a π - π interaction with naltrindole (Extended Data Fig. 2). This conformation of ECL3 is quite different from the one observed for the lower resolution *M. musculus* δ -OR structure⁸, which has an asparagine side chain instead of the Asp 290^{ECL3} seen in the human δ -OR. These high-resolution details of the binding pocket and ligand interactions in the human δ -OR orthosteric site provide an excellent framework for designing new δ -OR ligands¹⁴

and allosteric modulators¹⁵ with improved selectivity and functional profiles.

Unique features of the δ -OR sodium site

Evidence for the presence of a sodium ion in the allosteric site is similar to that observed in the high-resolution A_{2A}AR structure (PDB 4E1Y)¹⁶, including: (1) electron density showing coordination of the proposed sodium position by five oxygen atoms; (2) short distances observed between the ion and coordinating oxygens (~ 2.4 Å); and (3) calculations of ion valence (Supplementary Table 1). The cavity harbouring the allosteric sodium is formed by the side chains of 16 residues, 15 of which are highly conserved in class A GPCRs (Fig. 1 and Extended

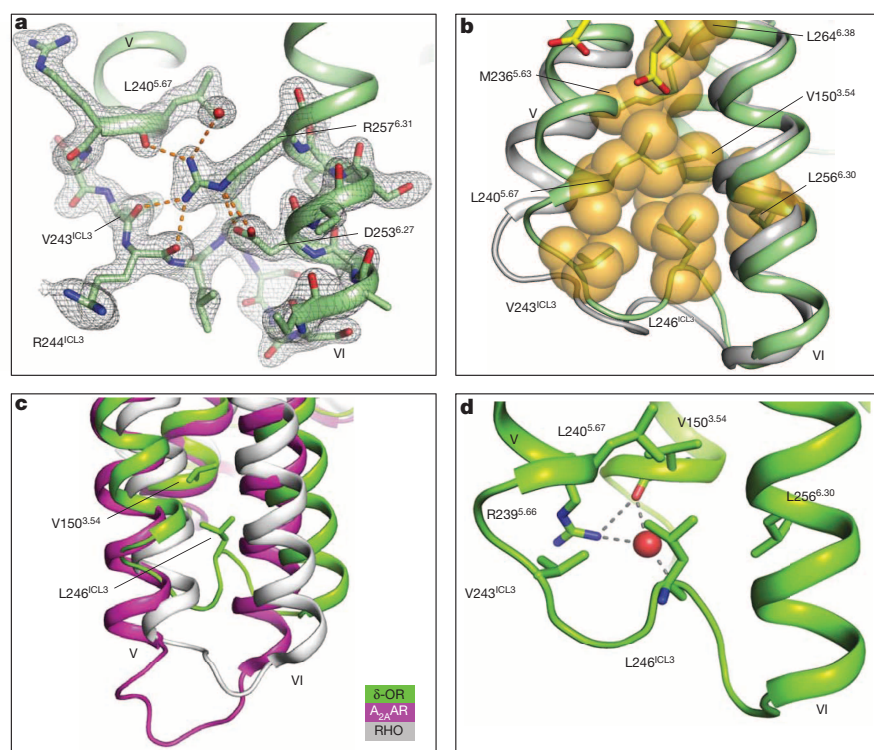


Figure 2 | Structure of the human δ -OR ICL3. **a**, $2mF_o - DF_c$ electron density map (grey mesh) of BRIL- δ OR(Δ N/ Δ C)-naltrindole ICL3 contoured at 1σ . Polar and ionic interactions of Arg 257^{6,31} are shown by orange dashed lines. **b**, ICL3 loop comparison between BRIL- δ OR(Δ N/ Δ C)-naltrindole (green) and NOP¹² (light grey) structures. Hydrophobic residues in the ICL3 hydrophobic cluster are shown as orange spheres and oleic acid (OLA) molecules are represented by yellow sticks. **c**, ICL3 comparison of BRIL- δ OR(Δ N/ Δ C)-naltrindole (green) with A_{2A}AR (magenta; PDB 3PWH) and rhodopsin (RHO, grey; PDB 3CAP). **d**, Details of the hydrogen bonds between Arg 239^{5,66}, Leu 246^{ICL3}, Val 150^{3,54} and water molecule (red sphere) are shown by grey dashed lines.

Data Fig. 3). Notably, the structure of BRIL- δ OR(Δ N/ Δ C)-naltrindole revealed that in addition to the highly conserved Asp 95^{2,50} and Ser 135^{3,39} side chains¹⁶, the sodium ion is directly coordinated by a non-conserved Asn 131^{3,35} side chain. Whereas Asn^{3,35} is conserved among opioid receptors (Extended Data Fig. 3b), the majority (~70%) of class A GPCRs have a hydrophobic residue in this position, and in the high-resolution A_{2A}AR structure the side chain of Leu 87^{3,35} is pointing towards the lipidic membrane¹⁶. By contrast, in the BRIL- δ OR(Δ N/ Δ C)-naltrindole structure the Asn 131^{3,35} side chain points into the sodium pocket, placing its oxygen (OD1) and nitrogen (ND2) atoms between the ion and the orthosteric pocket (Fig. 1). These exact atom positions are occupied by two water molecules in the allosteric sodium site of the A_{2A}AR structure (Extended Data Fig. 3a). In addition to the key role of the Asn 131^{3,35} side-chain OD1 atom in sodium coordination, the ND2 atom is hydrogen bonded to both side-chain OD1 and main-chain carbonyl atoms of Asp 128^{3,32} via a water molecule (Fig. 1); the latter residue occupies a central position deep in the orthosteric site and establishes a salt bridge with the nitrogen group of naltrindole. These interactions between the sodium ion, Asn 131^{3,35} and Asp 128^{3,32} establish an apparent axis of connectivity between orthosteric and allosteric regions on the receptor characterized in the inactive state. Altogether, the allosteric sodium of δ -OR is coordinated by five oxygen atoms, from Asp 95^{2,50}, Ser 135^{3,39} and Asn 131^{3,35} side chains and two structurally conserved water molecules, which comprises the first coordination shell for the sodium ion (Fig. 1 and Extended Data Figs 3, 4).

The second coordination shell of the sodium ion in the allosteric site is formed by the side chains of three residues (Trp 274^{6,48}, Asn 310^{7,45} and Asn 314^{7,49}) and two additional water molecules in contact with waters in the first shell (Fig. 1 and Extended Data Figs 3, 4). These conserved residues of the sodium pocket belong to two of the most well-known class A functional motifs: CW^{6,48}xP in helix VI and N^{7,49}PxxY in helix VII (x denotes any residue) (Fig. 1a), which have a critical role in GPCR activation processes¹⁷. As a whole, the cluster comprising the sodium ion and eight water molecules mediates extensive intrahelical hydrogen-bond networks between helices I, II, III, VI and VII in the

core of the receptor, when it is stabilized in an inactive state conformation. By contrast, the activated agonist-bound structure of A_{2A}AR reveals a sodium site that is collapsed by an inward movement of helix VII¹⁸, suggesting that rearrangements in this conserved sodium pocket have a key role in the activation of class A GPCRs².

Functional characterization of δ -OR

To correlate the structural data obtained using BRIL- δ OR(Δ N/ Δ C) with the wild-type δ -OR, we performed radioligand binding assays with opioid agonists and antagonists with wild-type δ -OR and BRIL- δ OR(Δ N/ Δ C) constructs expressed in HEK293 and Sf9 (*Spodoptera frugiperda*) cells, respectively. We found that the BRIL- δ OR(Δ N/ Δ C) and wild-type δ -OR displayed similar ligand-binding affinities (Extended Data Table 2) and, when both were expressed in HEK293 cells, they displayed similar functional coupling to G α_i -mediated signalling (Extended Data Fig. 5a, b). Consistent with classical studies performed on opioid receptors *in situ*^{5,19,20}, physiological concentrations of NaCl (140 mM) reduced the affinity of the δ -OR peptide agonist DADLE ([D-Ala², D-Leu⁵] enkephalin), which is structurally related to endogenous peptide agonists, at both wild-type δ -OR expressed in HEK293 cells and BRIL- δ OR(Δ N/ Δ C) expressed in Sf9 cells, while having minimal effects on antagonist-binding affinity (Extended Data Table 2). To confirm the specificity of the sodium site, we also examined the effects of other monovalent cations on the binding of ³H-DADLE in saturation binding assays. Among several monovalent cations tested, only sodium at physiological concentrations reduced ³H-DADLE binding (Extended Data Fig. 6).

Sodium modulates δ -OR ligand binding

Although the phenomenon of allosteric modulation of GPCR ligand affinity by sodium ions has been previously described for a number of class A GPCRs (for example, opioid, adrenergic, adenosine and dopamine receptors)^{21–24}, the nature of this allosteric effect, as well as sodium's affinity for the allosteric site is unknown. To quantify sodium's affinity at its allosteric binding site and to clarify the nature of the apparent negative cooperativity with respect to the peptide

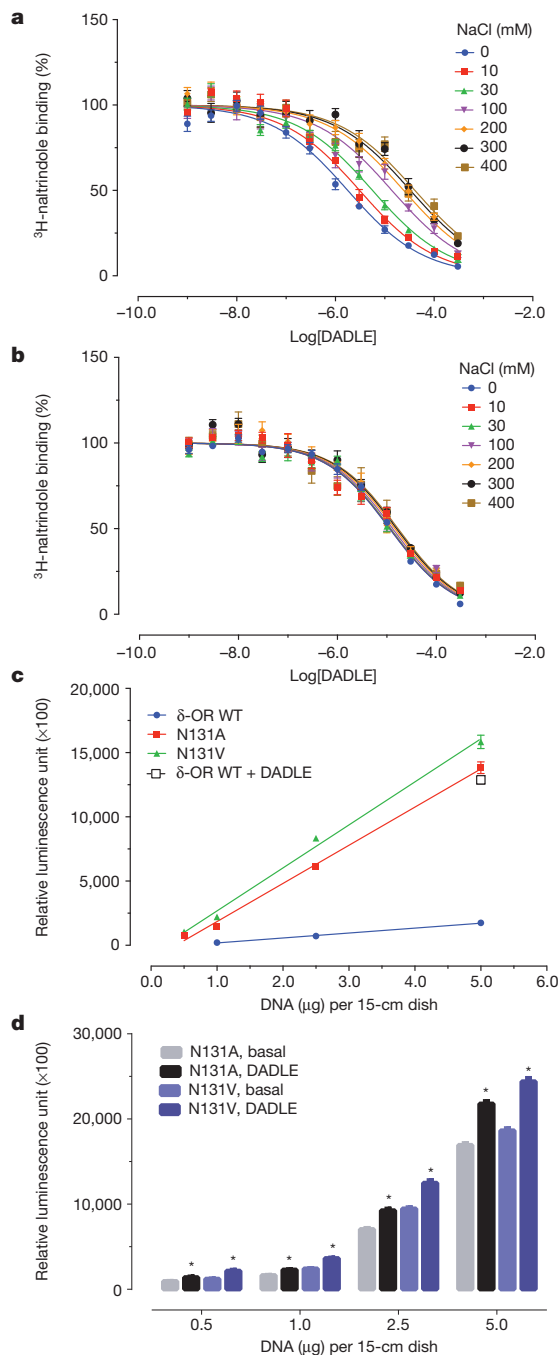


Figure 3 | Effect of sodium site mutations on sodium allosterism and β -arrestin constitutive activity. **a, b,** The effects of graded doses of sodium on DADLE affinity were measured at wild-type (**a**) and D95A mutant (**b**). Results were analysed using the allosteric model summarized in Table 1. **c,** The basal activity of Asn131^{3.35}Ala and Asn131^{3.35}Val is compared to wild type (WT) over a range of DNA dosages and their activity exceeded the one achieved with a saturating concentration of DADLE (10 μ M) at wild type (open square). Receptors were expressed at comparable levels (wild type, 226–758 fmol mg^{−1}; Asn131^{3.35}Ala, 77–553 fmol mg^{−1}; Asn131^{3.35}Val, 176–715 fmol mg^{−1}). Results represent average \pm s.e.m. from a minimum of 64 replicates from a representative assay. **d,** Mutants Asn131^{3.35}Ala and Asn131^{3.35}Val responded to DADLE (10 μ M) with a modest degree of stimulation ($*P < 0.01$ (*t*-test) versus no drug addition (basal)). Low-level expression was used allowing the detection of activation by a saturating concentration of DADLE.

agonist DADLE, we performed a series of radioligand binding assays in varying NaCl concentrations and then analysed the results using a standard allosteric model²⁵. Our studies revealed that sodium had

essentially the same affinity and produced a similar degree of negative cooperativity at both wild-type δ -OR expressed in HEK293 cells and the BRIL- δ OR(Δ N/ Δ C) expressed in Sf9 cells, although the negative cooperativity with DADLE was slightly decreased in the crystallized construct (Fig. 3a, Extended Data Fig. 5c, d and Table 1). These findings confirm that the crystallized construct maintains essentially the same capacity for sodium-dependent allosteric regulation as the wild-type δ -OR. Moreover, by revealing the relatively high sodium affinity to δ -OR ($K_B = 13.3$ mM), our studies demonstrate that at physiological sodium concentrations (140 mM) the sodium site is likely to be saturated.

Asn 131 regulates β -arrestin bias

Given the intimate relationships between residues coordinating allosteric sodium and motifs implicated in GPCR activation (for example, Asn 314^{7.49} and Tyr 318^{7.53} of the NP^{7.50}xxY motif; Fig. 1)⁴, we predicted that mutating selected sodium site residues could modulate δ -OR functionality. Accordingly, we performed functional studies on selected sodium site mutants (Figs 3 and 4, Table 1, Supplementary Table 2, Extended Data Table 3 and Extended Data Fig. 7) and discovered that mutating the sodium-anchoring δ -OR residue Asn 131^{3.35} into alanine or valine considerably enhanced constitutive activity for the β -arrestin pathway (Fig. 3c, d and Extended Data Fig. 7). Notably, the β -arrestin constitutive activity of Asn131^{3.35}Ala or Asn131^{3.35}Val mutants exceeded the activation levels of wild-type δ -OR achieved with a saturating concentration of the agonist DADLE (Fig. 3c), whereas G α_i protein basal activity remained unaffected (Extended Data Fig. 7). The Asn 131^{3.35} mutants increased receptor β -arrestin constitutive activity as well as DADLE-binding affinity, as compared with wild-type δ -OR, albeit DADLE efficacy for the β -arrestin pathway was greatly reduced (Supplementary Table 2 and Extended Data Table 3). Importantly, although the Asn131^{3.35}Ala mutation abolished the ‘sodium effect’, the receptor carrying the Asn131^{3.35}Val mutation retained sodium ion binding, although with lower affinity compared to wild-type receptor (Table 1).

The key differences in sodium ion binding affinity between the Asn 131^{3.35} mutants provided us with a model system to further clarify the role of sodium on canonical G α_i -protein-mediated signalling. Notably, the Asn131^{3.35}Ala mutant was inactive whereas the Asn131^{3.35}Val mutant, which partially retains the ‘sodium effect’, maintained G α_i activity, although with reduced agonist potency compared with the wild-type δ -OR (Fig. 4a and Supplementary Table 2). These data indicate that a complete disruption of the interactions between Asn 131^{3.35} and the sodium ion can induce high levels of constitutive activity at non-canonical β -arrestin signalling, while simultaneously abolishing canonical G-protein signalling, essentially inducing an ‘efficacy switch’ from the G α_i protein pathway to a β -arrestin pathway. These results reveal that the non-conserved residue Asn 131^{3.35} has an essential role in controlling both δ -OR functional selectivity and constitutive activity, probably through its structural role in coordinating the allosteric

Table 1 | Allosteric parameters for sodium at BRIL- δ OR(Δ N/ Δ C) and wild-type δ -OR

	Na ⁺ pK _B \pm s.e.m.	Na ⁺ K _B (mM)	p α \pm s.e.m.	α	Hill coefficient
δ -OR WT (HEK293)	1.88 \pm 0.10	13.3	1.93 \pm 0.34	0.012	0.56 \pm 0.16
D95A (HEK293)	NAE	NAE	NAE	NAE	NAE
D95N (HEK293)	NAE	NAE	NAE	NAE	NAE
N131A (HEK293)	NAE	NAE	NAE	NAE	NAE
N131V (HEK293)	1.11 \pm 0.15	77	1.61 \pm 0.86	0.025	0.61 \pm 0.02
BRIL- δ OR(Δ N/ Δ C) (Sf9)	1.79 \pm 0.11	15.9	0.86 \pm 0.05	0.138	0.79 \pm 0.02

Radioligand binding assays (Fig. 3a, b and Extended Data Fig. 5c, d) were analysed by the allosteric model (see ref. 25 for details). Here α defines the effect of allosteric ligand (in this case, sodium) on orthosteric ligand (in this case, DADLE); an $\alpha > 1$ indicates a positive effect thereby increasing binding affinity, whereas an $\alpha < 1$ indicates a negative effect thereby reducing binding affinity. For the Asp95^{2.50}Ala, Asp95^{2.50}Asn and Asn131^{3.35}Ala mutations, no allosteric effect of sodium was observed (see Fig. 3b for representative data with the Asp95^{2.50}Ala mutant). NAE, no allosteric effect of sodium observed; WT, wild type.

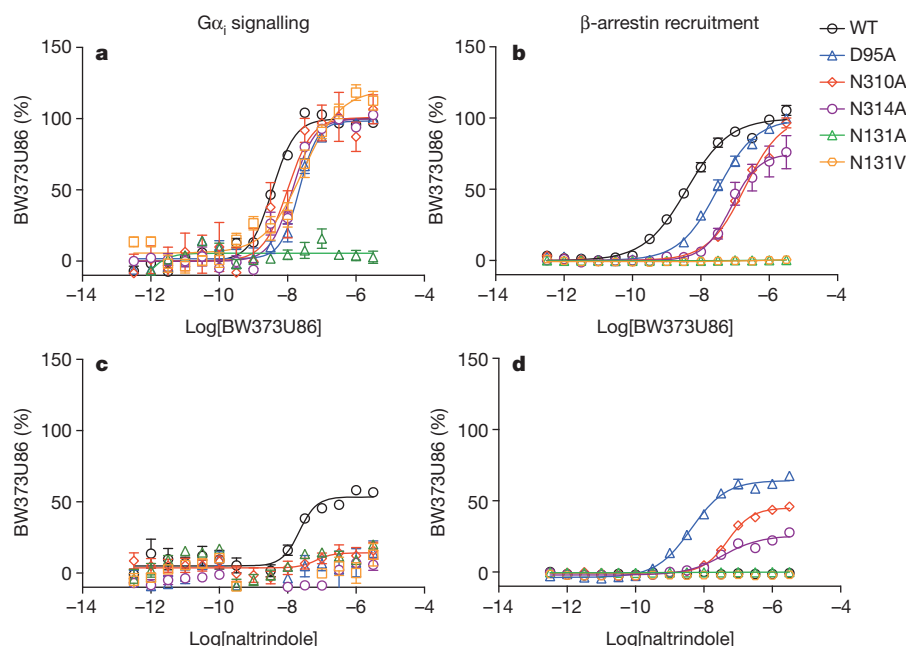


Figure 4 | Sodium-coordinating residues form an efficacy switch regulating biased signalling. Mutation of the sodium-anchoring residues Asp95^{2,50}, Asn310^{7,45} and Asn314^{7,49} promotes efficacy switching of the cyclopentene-containing antagonist naltrindole into a potent β -arrestin-biased agonist. **a–d**, Normalized concentration-responses of δ -OR-mediated $G\alpha_i$ signalling induced by BW373U86 (**a**) and naltrindole (**c**), and δ -OR-mediated β -arrestin

recruitment with BW373U86 (**b**) and naltrindole (**d**) were quantified as in Methods. Results represent average \pm s.e.m. of four independent experiments each in quadruplicate and are presented as percentage of activation by BW373U86. Receptors were all transfected with 15 μ g DNA revealing weak partial agonist activity of naltrindole at $G\alpha_i$ signalling as described previously^{29,30}.

sodium. We also compared basal $G\alpha_i$ protein activity for the sodium site δ -OR mutants (Extended Data Fig. 7) and found that the mutants have constitutive $G\alpha_i$ activity similar to the wild-type δ -OR, with a minor increase at Asn314^{7,49} Ala mutant and decrease at Asn131^{3,35} Ala and Asn131^{3,35} Val mutants. The strong β -arrestin-biased activity observed with Asn131^{3,35} Ala and Asn131^{3,35} Val could contribute to this small reduction by displacing $G\alpha_i$ protein.

Sodium-dependent opioid pharmacology

We next examined Asp95^{2,50}, another key sodium site residue, and found that mutation of this residue into either alanine or asparagine abolished the ‘sodium effect’ in radioligand binding assays (Fig. 3b and Table 1). The small molecule agonist BW373U86 has been previously described as a selective orthosteric δ -OR agonist for which binding to the receptor is minimally affected by sodium ions²⁶. Consequently, BW373U86 displays a low-to-moderate reduction in G-protein and β -arrestin signalling at sodium-anchored-residue mutants compared to wild type (Fig. 4a, b). On the other hand, the peptide agonist DADLE has a diminished potency for β -arrestin recruitment at the δ -OR mutant Asp95^{2,50} Ala and is inactive at both the Asn310^{7,45} Ala and the Asn314^{7,49} Ala mutants, while showing low-to-moderate reduction in G-protein-agonist efficacy (Supplementary Table 2). Importantly, we discovered that classical δ -OR ligands containing a cyclopentene functional group (naltrindole, naltriben and 7-benzylidenenaltrexone (BNTX)), which display no apparent agonist activity at the β -arrestin pathway with the wild-type δ -OR, gained potent β -arrestin-biased agonist activity at the Asp95^{2,50} Ala sodium site mutant (Fig. 4d and Supplementary Table 2). A similar transformative effect is observed when the sodium site residues Asn 310^{7,45} and Asn 314^{7,49} were mutated to alanine (Fig. 4d and Supplementary Table 2). The conversion of these antagonists/weak partial agonists into β -arrestin-biased agonists by mutations in the allosteric sodium site, together with the effects of Asn 131^{3,35} mutants described above, uncovered what we characterize as ‘efficacy switches’ within δ -OR. These efficacy switches are apparently distinct from those previously reported at which only G-protein signalling was enhanced^{27,28}.

The allosteric sodium-binding pocket described here in atomic detail is potentially an attractive drug discovery target. Although the binding of small molecule allosteric modulators in this highly conserved site is unlikely to have desired subtype selectivity, extension of selective orthosteric ligands into the sodium cavity may lead to bitopic compounds with new pharmacological properties, for example, inverse agonism or strong functional bias. The detailed crystal structure may also help to identify the binding site and shed light on the mode of action for the other types of positive allosteric modulator (PAM) compounds, such as those recently reported in ref. 15.

Our results reveal a profound and essential role for allosteric sodium-anchoring residues at specifying GPCR signal transduction and pharmacology. Mutation of sodium-anchoring residues within the allosteric site selectively modulates not only agonist binding, but also markedly changes GPCR functional activity by augmenting β -arrestin constitutive activity and introducing new patterns of biased signalling. In particular, these findings highlight the unexpectedly essential role for sodium-coordinating amino acids as efficacy switches for GPCR signalling.

METHODS SUMMARY

The BRIL- δ OR(Δ N/ Δ C) construct has its first 35 amino-terminal residues (Δ N 35) replaced with the thermostabilized apocytochrome *b*₅₆₂ RIL (M7W, H102I, R106L; BRIL) and 34 carboxy-terminal residues (Δ C 34) deleted and was expressed in *S. frugiperda* (Sf9) insect cells for structural and radioactive ligand-binding experiments. Sf9 membranes were solubilized in 0.75% (w/v) *n*-dodecyl- β -D-maltopyranoside and 0.15% (w/v) cholesteryl hemisuccinate in the presence of 25 μ M naltrindole, and purified by immobilized metal ion affinity chromatography. Receptor crystallization was performed using the lipidic cubic phase (LCP) method. The protein-LCP mixture contained 40% (w/w) protein solution, 54% (w/w) monoolein (Sigma) and 6% (w/w) cholesterol. Crystallization trials were performed using 40 nl protein-laden LCP overlaid with 0.8 μ l precipitant solution (31–34% (v/v) PEG 400, 0.095–0.12 M K/Na tartrate, 5% (v/v) ethylene glycol, 100 mM MES buffer, pH 6.1–6.2, and 1 mM naltrindole) at 20 °C. Crystallographic data were collected on the 23ID-D beamline (GM/CA CAT) of the Advanced Photon Source at the Argonne National Laboratory using a 20- μ m collimated minibeam. Data sets from 47 different crystals were merged for the final data set (Supplementary Table 1). cAMP assays were performed in HEK293T cells co-transfected with

human wild-type δ -OR or various mutants along with a split-luciferase-based cAMP biosensor (GloSensor; Promega). δ -OR β -arrestin-recruitment assays were performed using the Tango assay as described in Methods. ^3H -naltrindole-binding assays were performed using S9 membranes expressing the crystallized construct BRIL- δ OR($\Delta\text{N}/\Delta\text{C}$) or HEK293 T-cell membrane preparations transiently expressing wild-type or mutant δ -OR receptors.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 September; accepted 6 December 2013.

Published online 12 January 2014.

- Pasternak, G. W. Opioids and their receptors: are we there yet? *Neuropharmacology* **76**, 198–203 (2014).
- Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
- Wooten, D., Christopoulos, A. & Sexton, P. M. Emerging paradigms in GPCR allostery: implications for drug discovery. *Nature Rev. Drug Discov.* **12**, 630–644 (2013).
- Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Pert, C. B., Pasternak, G. & Snyder, S. H. Opiate agonists and antagonists discriminated by receptor binding in brain. *Science* **182**, 1359–1361 (1973).
- Cooper, D. M., Londos, C., Gill, D. L. & Rodbell, M. Opiate receptor-mediated inhibition of adenylate cyclase in rat striatal plasma membranes. *J. Neurochem.* **38**, 1164–1167 (1982).
- Portoghese, P. S., Sultana, M., Nagase, H. & Takemori, A. E. Application of the message-address concept in the design of highly potent and selective non-peptide δ opioid receptor antagonists. *J. Med. Chem.* **31**, 281–282 (1988).
- Granier, S. *et al.* Structure of the δ -opioid receptor bound to naltrindole. *Nature* **485**, 400–404 (2012).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Doré, A. S. *et al.* Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19**, 1283–1293 (2011).
- Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
- Thompson, A. A. *et al.* Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485**, 395–399 (2012).
- Bonner, G., Meng, F. & Akil, H. Selectivity of μ -opioid receptor determined by interfacial residues near third extracellular loop. *Eur. J. Pharmacol.* **403**, 37–44 (2000).
- Filizola, M. & Devi, L. A. Grand opening of structure-guided design for novel opioids. *Trends Pharmacol. Sci.* **34**, 6–12 (2013).
- Burford, N. T. *et al.* Discovery of positive allosteric modulators and silent allosteric modulators of the μ -opioid receptor. *Proc. Natl Acad. Sci. USA* **110**, 10830–10835 (2013).
- Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
- Audet, M. & Bouvier, M. Restructuring G-protein-coupled receptor activation. *Cell* **151**, 14–23 (2012).
- Xu, F. *et al.* Structure of an agonist-bound human A2A adenosine receptor. *Science* **332**, 322–327 (2011).
- Selley, D. E., Cao, C. C., Liu, Q. & Childers, S. R. Effects of sodium on agonist efficacy for G-protein activation in μ -opioid receptor-transfected CHO cells and rat thalamus. *Br. J. Pharmacol.* **130**, 987–996 (2000).
- Yabaluri, N. & Medzihradsky, F. Regulation of μ -opioid receptor in neural cells by extracellular sodium. *J. Neurochem.* **68**, 1053–1061 (1997).
- Horstman, D. A. *et al.* An aspartate conserved among G-protein receptors confers allosteric regulation of alpha 2-adrenergic receptors by sodium. *J. Biol. Chem.* **265**, 21590–21595 (1990).
- Costa, T., Lang, J., Gless, C. & Herz, A. Spontaneous association between opioid receptors and GTP-binding regulatory proteins in native membranes: specific regulation by antagonists and sodium ions. *Mol. Pharmacol.* **37**, 383–394 (1990).
- Gao, Z. G. & IJzerman, A. P. Allosteric modulation of A(2A) adenosine receptors by amiloride analogues and sodium ions. *Biochem. Pharmacol.* **60**, 669–676 (2000).
- Neve, K. A. Regulation of dopamine D2 receptors by sodium and pH. *Mol. Pharmacol.* **39**, 570–578 (1991).
- Christopoulos, A. & Kenakin, T. G protein-coupled receptor allostery and complexing. *Pharmacol. Rev.* **54**, 323–374 (2002).
- Childers, S. R., Fleming, L. M., Selley, D. E., McNutt, R. W. & Chang, K. J. BW373U86: a nonpeptidic delta-opioid agonist with novel receptor-G protein-mediated actions in rat brain membranes and neuroblastoma cells. *Mol. Pharmacol.* **44**, 827–834 (1993).
- Holst, B. *et al.* Identification of an efficacy switch region in the ghrelin receptor responsible for interchange between agonism and inverse agonism. *J. Biol. Chem.* **282**, 15799–15811 (2007).
- Steen, A. *et al.* Biased and constitutive signaling in the CC-chemokine receptor CCR5 by manipulating the interface between transmembrane helices 6 and 7. *J. Biol. Chem.* **288**, 12511–12521 (2013).
- Szekeres, P. G. & Traynor, J. R. Delta opioid modulation of the binding of guanosine 5'-O-(3-[35S]thio)triphosphate to NG108–15 cell membranes: characterization of agonist and inverse agonist effects. *J. Pharmacol. Exp. Ther.* **283**, 1276–1284 (1997).
- Liu, J. G. & Prather, P. L. Chronic agonist treatment converts antagonists into inverse agonists at δ -opioid receptors. *J. Pharmacol. Exp. Ther.* **302**, 1070–1079 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Institutes of Health Common Fund grant P50 GM073197 for technology development (V.C. and R.C.S.), PSI: Biology grant U54 GM094618 for biological studies and structure production (target GPCR-39) (V.K., V.C. and R.C.S.), R01 DA017204 and the NIMH Psychoactive Drug Screening Program (P.G., X.-P.H., B.L.R.) and the Michael Hooker Chair for Protein Therapeutics and Translational Proteomics to B.L.R. We thank J. Velasquez for help with molecular biology, T. Trinh and M. Chu for help with baculovirus expression, G.W. Han for help with structure analysis and quality control review, E. Abola for help with sodium site analysis, A. Walker for assistance with manuscript preparation and J. Smith, R. Fischetti and N. Sanishvili for assistance in development and use of the minibeam and beamtime at beamline 23-ID at the Advanced Photon Source, which is supported by National Cancer Institute grant Y1-CO-1020 and National Institute of General Medical Sciences grant Y1-GM-1104.

Author Contributions G.F. designed, optimized and purified δ -OR receptor constructs for structural studies, crystallized the receptor in LCP, collected and processed diffraction data, determined the structure, analysed the data and wrote the paper. P.M.G. performed mutagenesis and signalling studies, analysed the data and wrote the paper. X.-P.H. performed ligand binding and signalling studies, analysed the data and wrote the paper. V.K. analysed the data and wrote the paper. A.A.T. designed and cloned initial δ -OR constructs. V.C. analysed the data and wrote the paper. B.L.R. supervised the pharmacology and mutagenesis studies, analysed the data and wrote the paper. R.C.S. was responsible for the overall project strategy and management, analysed the data and wrote the paper.

Author Information The coordinates and the structure factors have been deposited in the Protein Data Bank under accession code 4N6H. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu) or B.L.R. (bryan_roth@med.unc.edu).

METHODS

Cloning, expression and purification. The wild-type human δ -OR gene (*OPRD1*; UniProt accession P41143) was synthesized by DNA2.0 with codon optimization for expression in *S. frugiperda* (*Sf9*), and then cloned into a modified pFastBac1 vector (Invitrogen) containing an expression cassette with a haemagglutinin signal sequence followed by a Flag tag, a $10 \times$ His tag and a tobacco etch virus (TEV) protease recognition site at the N terminus. Thirty-four amino acids were deleted from the C terminus (residues 339–372), and 35 residues of the N terminus (residues 1–35) of δ -OR were replaced with the thermostabilized apocytocrome b_{562} RIL from *Escherichia coli* (M7W, H102I and R106L) (BRIL)³¹ protein using splicing by overlap extension PCR³². A Pro37Ser mutation was introduced in the N terminus of δ -OR to facilitate crystallization. Recombinant baculoviruses were generated using the Bac-to-Bac system (Invitrogen) and were used to infect *Sf9* insect cells at a density of 2×10^6 cells ml^{-1} at a multiplicity of infection of 5 as described previously³³. Infected cells were grown at 27 °C for 48 h before being collected, and the cell pellets were stored at -80 °C.

Insect cell membranes were disrupted as described previously³⁴. In brief, cell pellets were homogenized by douncing in a hypotonic buffer containing 10 mM HEPES, pH 7.5, 10 mM MgCl_2 , 20 mM KCl and EDTA-free complete protease inhibitor cocktail tablets (Roche). Washing of the membranes was performed by repeated dounce homogenization and centrifugation in the same hypotonic buffer (once more), followed by high osmotic buffer containing 1.0 M NaCl, 10 mM HEPES, pH 7.5, 10 mM MgCl_2 , 20 mM KCl and EDTA-free complete protease inhibitor cocktail tablets (three times). Purified membranes were re-suspended in 10 mM HEPES, pH 7.5, 10 mM MgCl_2 , 20 mM KCl and 30% (v/v) glycerol, flash frozen with liquid nitrogen, and stored at -80 °C.

Before receptor purification membranes were thawed and washed once again in a buffer containing 1.0 M NaCl, 10 mM HEPES, pH 7.5, 10 mM MgCl_2 , 20 mM KCl. Washed membranes were re-suspended in buffer containing 50 μM naltrindole (Tocris), 2 mg ml^{-1} iodoacetamide (Sigma), 500 mM NaCl, 50 mM HEPES, pH 7.5, and incubated at 4 °C for 1 h before solubilization. The membranes were then solubilized in 50 mM HEPES, pH 7.5, 500 mM NaCl, 0.75% (w/v) *n*-dodecyl- β -D-maltopyranoside (DDM; Anatrace), 0.15% (w/v) cholesteryl hemisuccinate (CHS; Sigma) and 25 μM naltrindole for 3 h at 4 °C. The supernatant was isolated by centrifugation at 160,000g for 45 min, and incubated in 20 mM buffered imidazole (pH 7.5), 0.7 M NaCl with 1 ml of TALON IMAC resin (Clontech) overnight at 4 °C. After binding, the resin was washed with 15 column volumes of wash buffer I (50 mM HEPES, pH 7.5, 800 mM NaCl, 10% (v/v) glycerol, 0.1% (w/v) DDM, 0.02% (w/v) CHS, 10 mM ATP, 10 mM MgCl_2 and 50 μM naltrindole), followed by 10 column volumes of wash buffer II (50 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.02% (w/v) DDM, 0.004% (w/v) CHS, 50 mM imidazole and 50 μM naltrindole). The protein was then eluted by 5 column volumes of elution buffer (25 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.02% (w/v) DDM, 0.004% (w/v) CHS, 250 mM imidazole and 100 μM naltrindole). PD miniTrap G-25 column (GE Healthcare) was used to remove imidazole. The protein was then treated overnight with His-tagged TEV protease to cleave the N-terminal His tag and Flag tag. TEV protease and the cleaved N-terminal fragment were removed by TALON IMAC resin incubation for 1 h at 4 °C. Purified receptor was concentrated to 20 mg ml^{-1} with a 100 kDa molecular weight cutoff Vivaspin centrifuge concentrator (GE Healthcare). Protein purity and monodispersity were tested by SDS-PAGE and analytical size-exclusion chromatography (aSEC). Typically, the protein purity exceeded 95%, and the aSEC profile showed a single peak, indicative of receptor monodispersity.

Crystallization. Receptor samples in complex with naltrindole were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid using a mechanical syringe mixer³⁵. The protein-LCP mixture contained 40% (w/w) protein solution, 54% (w/w) monoolein (Sigma) and 6% (w/w) cholesterol (AvantiPolar Lipids). Crystallization trials were performed in 96-well glass sandwich plates³⁶ (Marienfeld) by an NT8-LCP crystallization robot (Formulatrix) using 40 nl protein-laden LCP overlaid with 0.8 μl precipitant solution in each well, and sealed with a glass coverslip. Protein reconstitution in LCP and crystallization trials was carried out at room temperature (~ 20 – 23 °C). The crystallization plates were stored and imaged in an incubator/imager (RockImager 1000, Formulatrix) at 20 °C. Diffraction quality crystals of an average size of $50 \times 10 \times 3$ μm were obtained within ~ 10 days in 31–34% (v/v) PEG 400, 0.095–0.12 M K/Na tartrate, 5% (v/v) ethylene glycol, 100 mM MES buffer, pH 6.1–6.2, and 1 mM naltrindole. Crystals were collected directly from LCP using 50 μM MiTeGen micromounts and immediately flash frozen in liquid nitrogen.

X-ray data collection and processing. Crystallographic data were collected on the 23ID-D beamline (GM/CA CAT) of the Advanced Photon Source at the Argonne National Laboratory using a 20- μm collimated minibeam at a wavelength of 1.0330 Å and a MarMosaic 300 detector. To reduce radiation damage crystals were translated to a fresh position, if possible, or replaced after collecting

10 frames at 1-s exposure and 1.0° oscillation with an unattenuated beam. Data sets from 47 different crystals were integrated, scaled and merged together using HKL2000³⁷ (Extended Data Table 1).

Structure determination and refinement. Initial molecular replacement solution was obtained by PHASER³⁸ in the CCP4 suite, using the 3.4 Å δ -OR receptor structure (PDB 4EJ4) with deleted T4L, and BRIL from A_{2A}AR (PDB 4E1Y) as independent search models. The resulting BRIL- δ OR(Δ N/ Δ C) model was refined by manually building in the excessive $2F_o - F_c$ density and by repetitive cycling between COOT³⁹, REFMAC5⁴⁰, and simulated annealing using PHENIX⁴¹ until convergence. Ten translation, libration and screw-rotation atomic displacement (TLS) groups were used throughout refinement. The elongated electron density tubes near the protein hydrophobic surface were modelled as oleic acids, with the exception of the few that were better fit with monooleins, the major lipid component used for crystallization. The data collection and refinement statistics are shown in Extended Data Table 1.

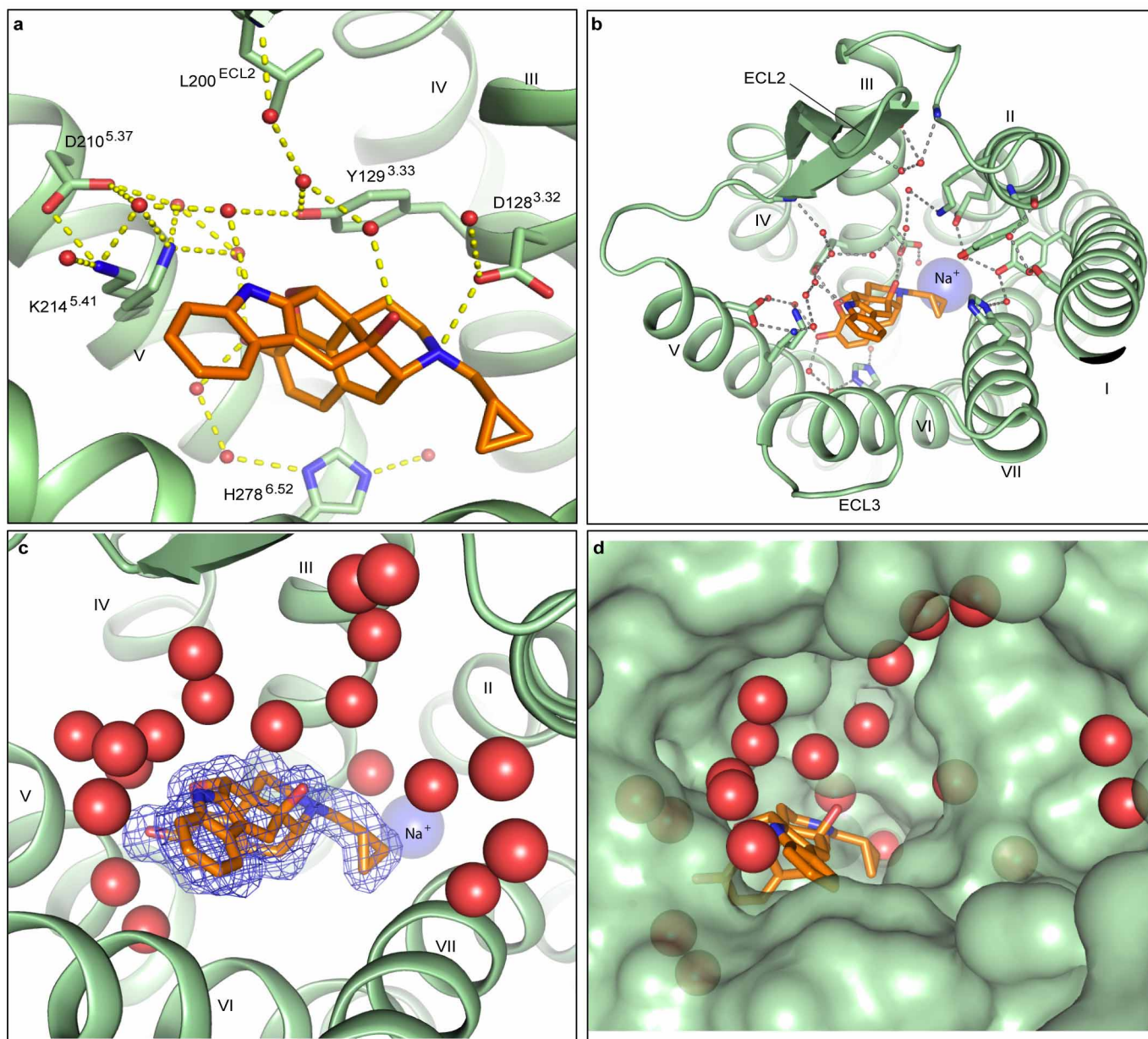
δ -OR β -arrestin-recruitment assay. Assays were performed using modifications of the original Tango assay⁴² as described previously^{43,44}. In brief, δ -OR constructs were codon optimized for mammalian expression and synthesized by Blue Heron Biotech, and inserted into the modular vector system (described in ref. 42) with all constructs confirmed by automated double-stranded DNA sequencing. HTLA cells (provided by R. Axel) were transfected by the calcium phosphate precipitation method. The next day, the cells were plated in DMEM supplemented with 1–2% dialysed FBS into Poly-L-Lys-coated 384-well white clear bottom cell culture plates at a density of 15,000 cells per well in a total volume of 50 μl . The following day, ligand solutions were prepared in filtered assay buffer (20 mM HEPES, $1 \times$ Hanks' balanced salt solution (HBSS), pH 7.40) at 3.5 \times and added to cells (20 μl per well) for overnight incubation (16–20 h). The next day, media and drug solutions were removed and 20 μl per well of Bright-Glo reagent (Promega, 1:20 dilution in assay buffer) was added. Plates were incubated for 20 min at room temperature in the dark before being counting using a TriLux luminescence reader (1 s per well). To determine basal and constitutive activity of wild-type δ -OR and N131A or N131V mutant δ -ORs, HTLA cells were transfected as above at 0.5–5 μg DNA per 15-cm dish. An aliquot of the transfected cells was plated and incubated for overnight without agonist stimulation, and luminescence was determined as above. The remainders of the transfected cells were subjected to membrane preparation and binding assays to estimate receptor expression levels with ³H-naltrindole (see below). Data were subjected to nonlinear least-squares regression analysis using the sigmoidal dose-response function provided in GraphPad Prism 5.0 or 6.0. Data of four independent experiments ($n = 4$) performed in quadruplicate are presented as relative luminescence unit (Fig. 3) or as percentage of BW373U86 (Fig. 4).

cAMP assays. HEK293T (ATCC CRL-11268, Manassas) cells were co-transfected with human δ -OR or various mutants along with a split-luciferase-based cAMP biosensor (GloSensor; Promega) and assays performed in a manner similar to previous studies with G α_i -coupled receptors⁴⁵. Transfected cells were plated into Poly-L-Lys-coated 384-well white clear bottom cell culture plates with DMEM + 1% dialysed FBS at a density of 15,000 cells per 40 μl per well overnight. On the day of the assay, cells were removed of culture medium and received 20 μl per well of assay buffer (20 mM HEPES, $1 \times$ HBSS, pH 7.40) followed by addition of 10 μl of $3 \times$ drug solutions for 15 min at room temperature. To measure agonist activity for G α_i -coupled receptors, 10 μl luciferin (4 mM final concentration) supplemented with isoproterenol (400 nM final concentration) were added to activate G α_s via endogenous β_2 -adrenergic receptors, and luminescence intensity was quantified 15 min later. Data were normalized compared to the agonist BW373U86 and regressed using the sigmoidal dose-response function. Data of four independent experiments ($n = 4$) conducted in quadruplicate are presented as percentage of BW373U86 (Fig. 4). Analyses were performed using the software GraphPad Prism.

Radio ligand binding assays. ³H-naltrindole binding assays were performed using *Sf9* membrane fractions expressing the crystallization construct BRIL- δ OR(Δ N/ Δ C) or HEK293 T membrane preparations transiently expressing wild-type or mutant δ -OR receptors. HEK293 T cells were transfected to make membranes and binding assays were set up in 96-well plates as described previously⁴⁶. To examine the effects of sodium on ligand binding at δ -OR receptors, all binding assays were conducted in the δ -OR binding buffer (50 mM Tris HCl, 2 mM EDTA, pH 7.40) in the absence or presence of NaCl at designed concentrations. Saturation binding assays with 0.2–30 nM ³H-naltrindole in δ -OR binding buffer were performed to determine equilibrium dissociation constant (K_d), whereas 10 μM naltrindole was used to define nonspecific binding. To quantify the allosteric effects of sodium ions, test ligands (for example, DADLE) in a series of 11 concentrations (ranging from 0.1 nM–300 μM) were incubated with a fixed concentration of ³H-naltrindole in the absence and presence of increasing concentrations of sodium chloride (0.3 mM–1 M). Reactions (either saturation or competition binding) were incubated

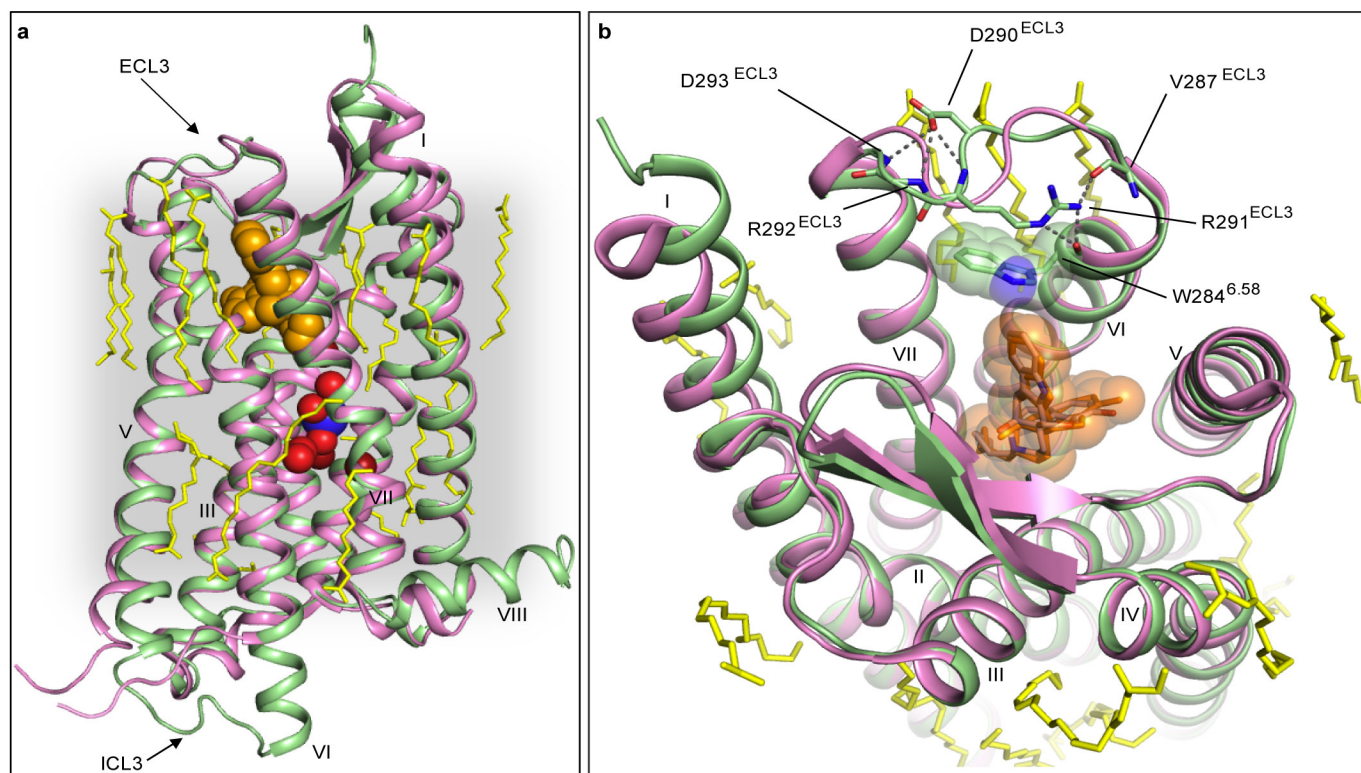
for 2 h at room temperature in the dark, and terminated by rapid vacuum filtration onto chilled 0.3% PEI-soaked GF/A filters followed by three quick washes with cold washing buffer (50 mM Tris HCl, pH 7.40) and quantified as described previously⁴⁶. Results (with or without normalization) were analysed using GraphPad Prism using one-site, two-site, or allosteric IC₅₀ shift models where indicated.

31. Chu, R. *et al.* Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *J. Mol. Biol.* **323**, 253–262 (2002).
32. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* **2**, 924–932 (2007).
33. Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).
34. Wacker, D. *et al.* Conserved binding mode of human β_2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* **132**, 11443–11445 (2010).
35. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
36. Cherezov, V., Peddi, A., Muthusubramanian, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60**, 1795–1807 (2004).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
41. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
42. Barnea, G. *et al.* The genetic design of signaling cascades to record receptor activation. *Proc. Natl Acad. Sci. USA* **105**, 64–69 (2008).
43. Allen, J. A. *et al.* Discovery of β -arrestin-biased dopamine D₂ ligands for probing signal transduction pathways essential for antipsychotic efficacy. *Proc. Natl Acad. Sci. USA* **108**, 18488–18493 (2011).
44. Carlsson, J. *et al.* Ligand discovery from a dopamine D₃ receptor homology model and crystal structure. *Nature Chem. Biol.* **7**, 769–778 (2011).
45. Wacker, D. *et al.* Structural features for functional selectivity at serotonin receptors. *Science* **340**, 615–619 (2013).
46. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215–220 (2012).



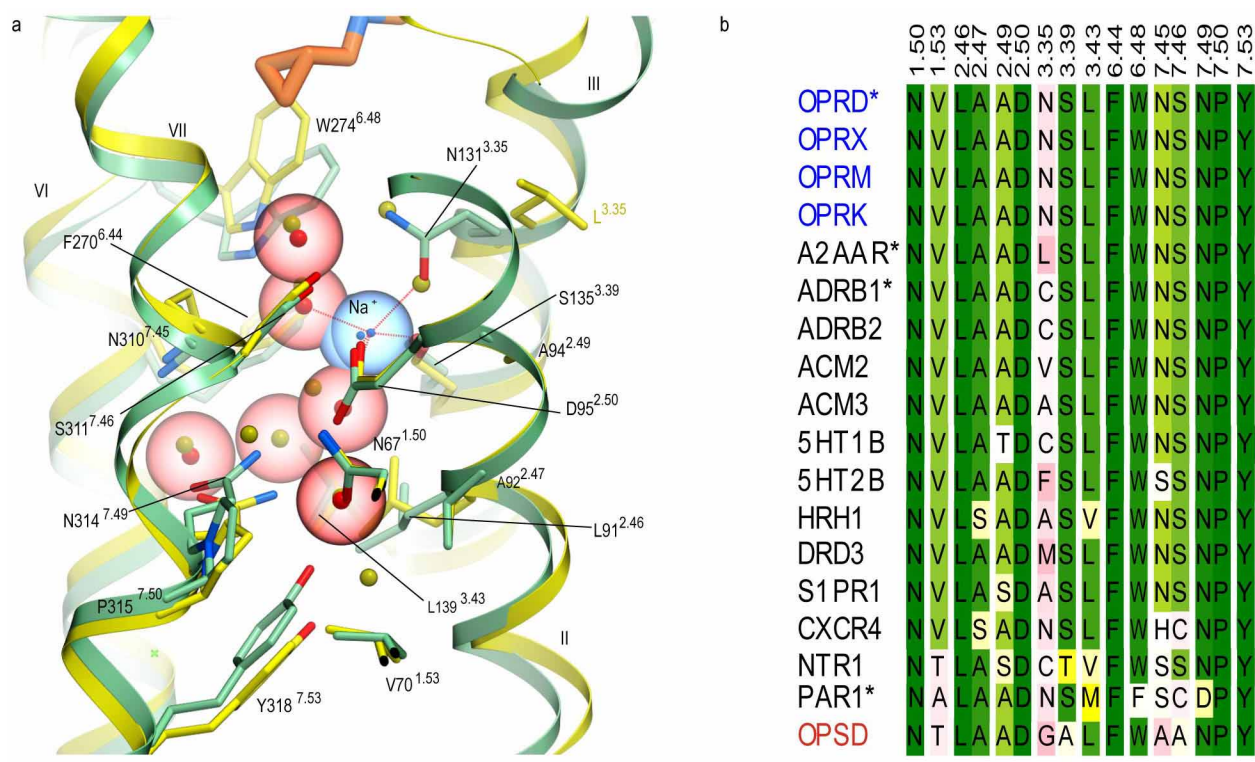
Extended Data Figure 1 | Interactions in the orthosteric ligand-binding site of BRIL-δOR(ΔN/ΔC)-naltrindole. **a**, 'Close up' view of the hydrogen-bond interaction network (yellow dotted lines) between the ligand naltrindole (orange sticks) and receptor (green cartoon; side chains shown as sticks) showing water (red spheres)-mediated interactions. **b**, Overview of water-mediated interactions (grey dotted lines) in the orthosteric pocket of BRIL-δOR(ΔN/ΔC)-naltrindole. **c**, $mF_o - DF_c$ 'omit' electron density map (blue

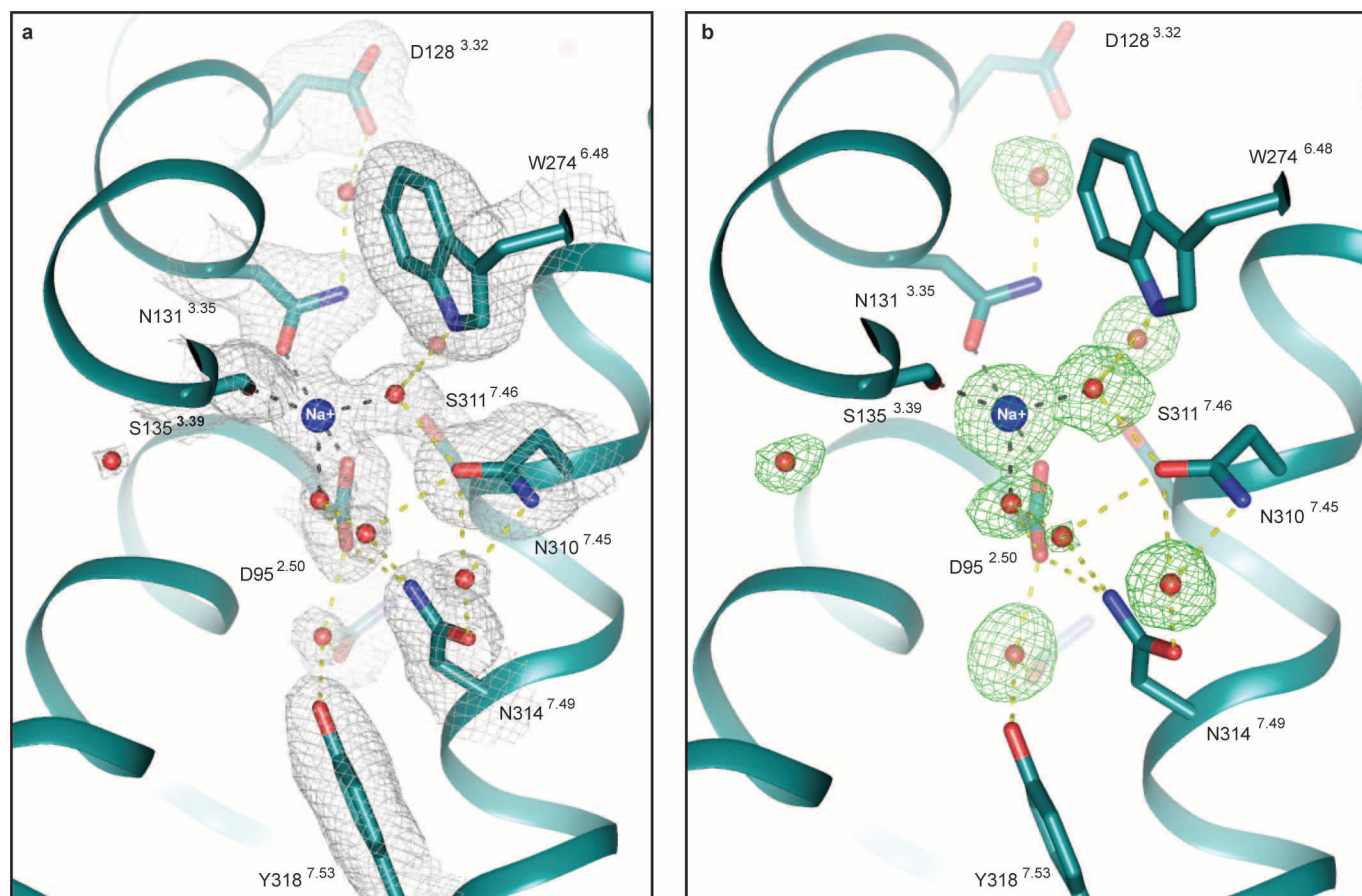
mesh) around naltrindole (orange sticks) after simulated annealing in PHENIX⁴¹ (5,000 K; contoured at 3σ); all water molecules surrounding the ligand in the orthosteric site are shown (red spheres). **d**, Surface representation of the BRIL-δOR(ΔN/ΔC)-naltrindole (green surface) showing the position of naltrindole (orange sticks) and the water cluster (red spheres) in the orthosteric site.



Extended Data Figure 2 | Superposition of *M. musculus* and human δ -OR receptor structures. **a**, Comparison of the human BRIL- δ OR(Δ N/ Δ C)-naltrindole structure (green) with the previously determined 3.4 Å crystal structure of the *M. musculus* δ OR-T4L (magenta; PDB 4EJ4). The ligand naltrindole in the BRIL- δ OR(Δ N/ Δ C)-naltrindole structure is shown as orange spheres and is omitted in the 4EJ4 structure. Water molecules and the sodium ion in the allosteric site of the BRIL- δ OR(Δ N/ Δ C)-naltrindole structure are shown as red and blue spheres, respectively. The grey box represents the approximate regions of the receptor 7TM domain embedded in

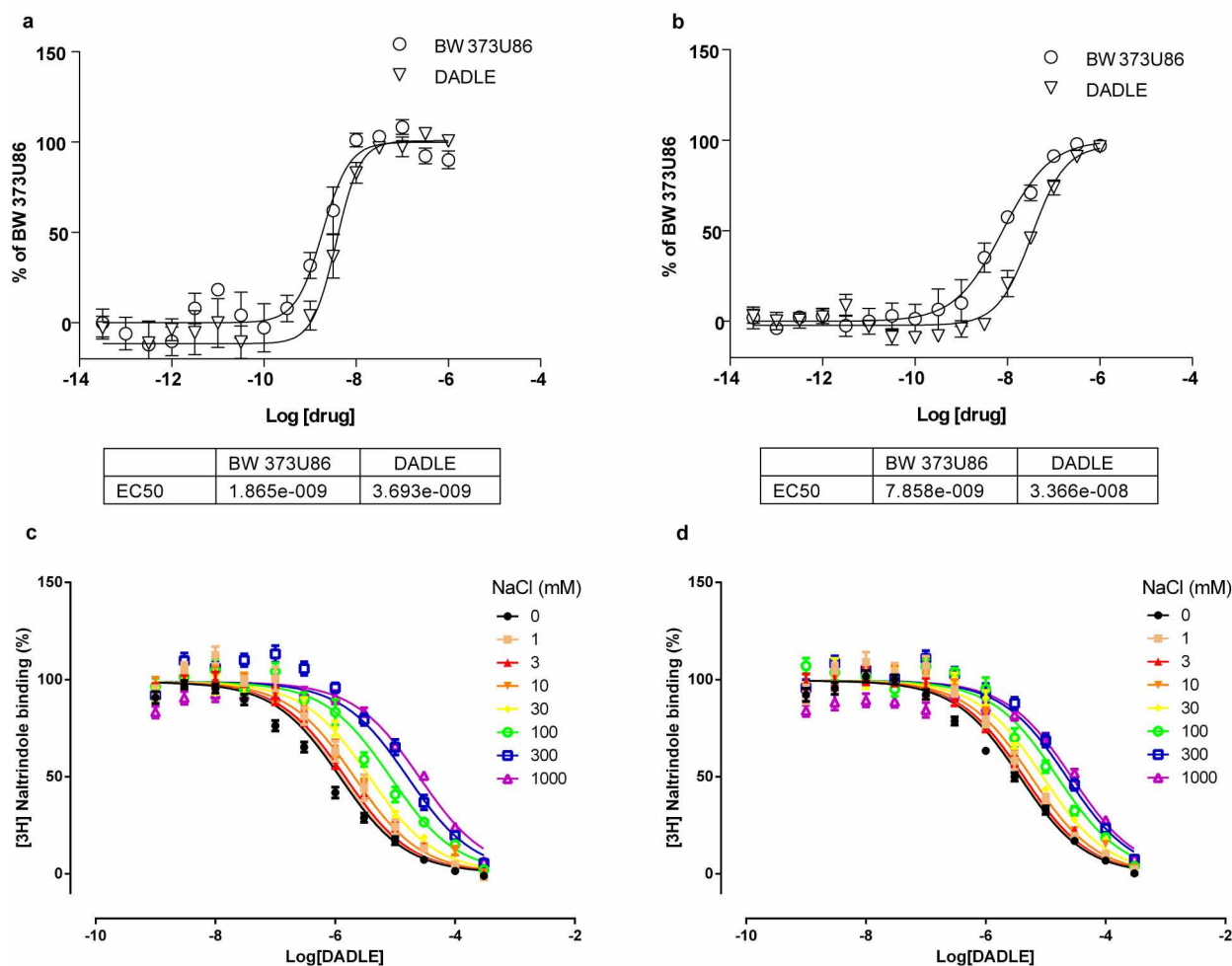
the lipidic bilayer; yellow sticks represent lipids. ICL3 fusion protein in 4EJ4 (T4L) and N-terminal BRIL fusion protein in BRIL- δ OR(Δ N/ Δ C)-naltrindole structures are omitted. **b**, ECL3 conformation of human δ -OR and comparison with mouse δ OR-T4L ECL3 structure. BRIL- δ OR(Δ N/ Δ C)-naltrindole is shown as green cartoon and side chains are depicted as green sticks. The ligand naltrindole is represented by orange carbon sticks and transparent spheres (BRIL- δ OR(Δ N/ Δ C)-naltrindole structure) and magenta carbon sticks (4EJ4 structure).





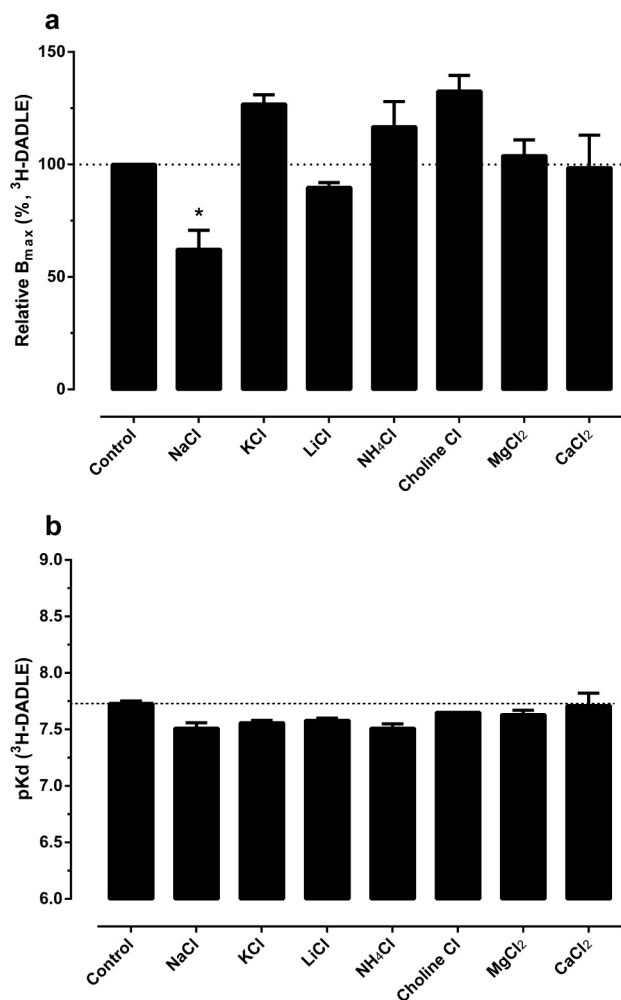
Extended Data Figure 4 | Electron density maps around the δ -OR sodium-binding site. **a**, $2mF_o - DF_c$ electron density map (grey mesh) of receptor residues (cyan sticks) around the sodium ion (blue sphere)-binding site contoured at 1σ . Black dashed lines represent interactions between the sodium ion and the atoms in direct contact with the receptor residues or water molecules in the first hydration shell of the ion. Yellow dashed lines show the

hydrogen-bond network of water molecules and receptor residues. These polar interactions in the core of the 7TM bundle of the receptor establish an axis of connectivity between the orthosteric binding site, allosteric sodium site and residues in the intracellular side of the receptor. **b**, $mF_o - DF_c$ 'omit' electron density map of sodium ion and coordinating water molecules after simulated annealing in PHENIX⁴¹ (5,000 K; contoured at 3σ).

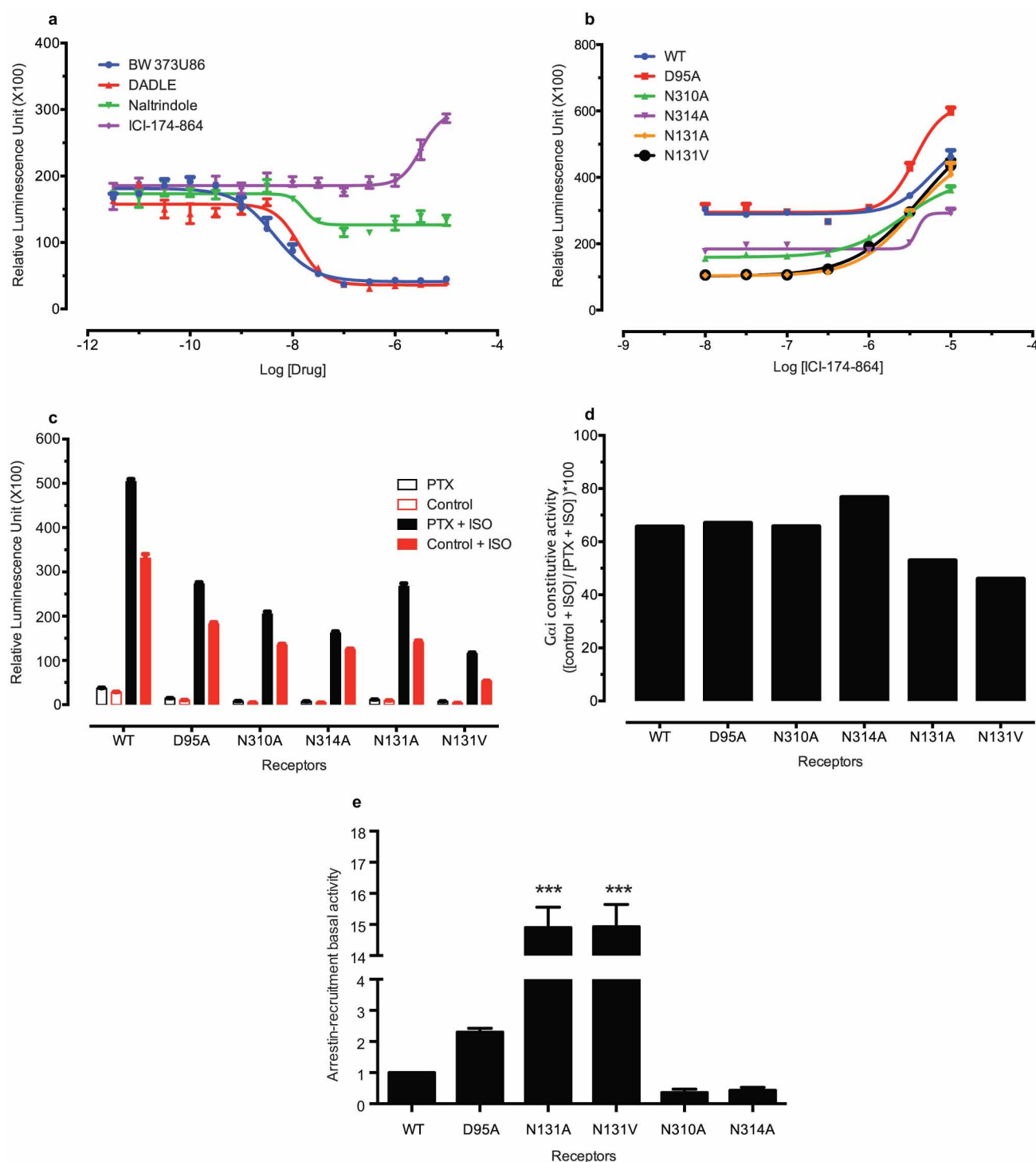


Extended Data Figure 5 | Functional characterization of wild-type δ -OR and the engineered BRIL- δ OR(Δ N/ Δ C) construct used for crystallization. **a, b**, G-protein signalling characterization for wild-type (**a**) and BRIL- δ OR(Δ N/ Δ C) (**b**) is shown. Allosteric inhibitory effect of sodium binding to 3 H-naltrindole-labelled wild-type δ -OR expressed in HEK293 cells

(positive control) (**c**) and the BRIL- δ OR(Δ N/ Δ C) construct used for crystallization and expressed in *Sf9* cells (**d**). Ligand-binding data (**d**) were analysed using the allosteric model, and the results are shown in Table 1. Data represent the mean of four independent experiments each in quadruplicate.



Extended Data Figure 6 | Effect of different cations on the binding of DADLE to BRIL- δ OR(Δ N/ Δ C). **a, b,** The effects of different cations on both the total number of binding sites (relative B_{max} in **a**) and affinity (pK_d in **b**) of $^3\text{H-DADLE}$ at the crystallized δ -OR construct expressed in *Sf9* cells. $^3\text{H-DADLE}$ saturation binding assays were conducted with control binding buffer (50 mM Tris HCl, pH 7.40) in the absence (control) and presence of 100 mM of the indicated chloride salts. Results were analysed in Prism to obtain B_{max} and K_d values. The B_{max} values were then normalized to control (100%) in each assay. Values represent mean \pm s.e.m. from three independent assays, each in triplicate. * $P < 0.05$, one-way analysis of variance (ANOVA).



Extended Data Figure 7 | Effect of sodium site mutations on basal $G\alpha_i$ and β -arrestin signalling. HEK293 cells were transfected with 10 μ g of receptor DNA and 10 μ g of GloSensor per 15-cm dish and wild-type and mutant δ -ORs were assayed for $G\alpha_i$ signalling as described in Methods. **a**, Concentration-responses of wild-type- δ -OR-mediated $G\alpha_i$ signalling induced by the agonists BW373U86 and DADLE, the antagonist† naltrindole, and the inverse agonist ICI-174-864. **b**, Concentration-responses of δ -OR wild-type- and mutant-mediated $G\alpha_i$ signalling induced by the inverse agonist ICI-174-864. **c**, **d**, Cells were treated with 100 ng ml⁻¹ pertussis toxin (PTX) for 12 h, depleting all $G\alpha_i$ constitutive activity. **c**, Non-normalized basal cAMP level of the δ -OR wild type and mutants in the presence or absence of PTX and isoproterenol (ISO). **d**, Net $G\alpha_i$ constitutive activity was calculated using

formula: cAMP level (RLU) from control + ISO / cAMP level (RLU) from PTX + ISO × 100. The data in **a** and **b** represent the mean \pm s.e.m. of at least four different experiments each in quadruplicate. The data in **c** and **d** represent the mean \pm s.e.m. of 32 wells ($n = 32$). †Weak partial agonist activity of naltrindole at $G\alpha_i$ signalling was described previously^{29,30}. **e**, Effect of sodium site mutations on basal β -arrestin signalling. Wild-type and mutant δ -ORs were assayed for β -arrestin recruitment as described in Methods. HTLA cells were transfected with 15 μ g of receptor DNA per 15-cm dish and basal activity measured 48 h later. Data are presented relative to wild-type receptor using the means \pm s.e.m. of at least four different experiments each in quadruplicate. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; one-way ANOVA.

Extended Data Table 1 | Data collection and refinement statistics

Structure		BRIL-δOR($\Delta N/\Delta C$)-naltrindole
Data Collection		(APS GM/CA CAT 23ID-D, 20 μm beam)
Number of crystals		47
Space group		C2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)		86.23, 72.85, 84.41
α , β , γ (°)		90.0, 107.4, 90.0
Number of reflections measured		204,142
Number of unique reflections		45,488
Resolution (Å)*		40.0-1.80 (1.86-1.80)
R _{merge} (%)		10.4 (82.2)
Mean <i>I</i> / σ <i>I</i>		16.4 (2.3)
Completeness		98.5 (98.5)
Redundancy		4.5 (4.3)
Refinement		
Resolution		29.7-1.80
Number of reflection (test set)		43,157 (2,303)
R _{work} /R _{free} (%)		17.1/19.0
Number of atoms		
Protein		3,140
Naltrindole		31
Sodium ion		1
Lipids and other		410
Overall B values (Å²)		
δ OR($\Delta N/\Delta C$)		34.3
BRIL		49.8
Naltrindole		26.2
Na ⁺		26.6
Lipids and other		60.9
Rmsd		
Bond lengths (Å)		0.012
Bond angles (°)		1.4
Ramachandran plot statistics (%)†		
Most favoured regions		99.5
Additionally allowed regions		0.5
Disallowed regions		0

$R_{\text{merge}} = \sum_{\text{hkl}} |I(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \langle I(\text{hkl}) \rangle$, where $\langle I(\text{hkl}) \rangle$ is the mean of the symmetry equivalent reflections of $I(\text{hkl})$.

* Highest resolution shell is shown in parentheses.

† As defined in MolProbity.

Extended Data Table 2 | Analysis of the effects of sodium ions on the binding of δ -OR ligands

Ligand type and name		HEK 293 WT δ -OR		Sf9 BRIL- δ OR(Δ N/ Δ C)	
		$pK_i \pm \text{SEM} (K_i, \text{nM})$		$pK_i \pm \text{SEM} (K_i, \text{nM})$	
		0 NaCl	140mM NaCl	0 NaCl	140mM NaCl
Small molecule	(-) Bremazocine	7.46 ± 0.06 (35)	7.33 ± 0.05 (47)	7.88 ± 0.11 (13)	7.84 ± 0.12 (14)
	(-) Cyclazocine	7.01 ± 0.07 (97)	6.81 ± 0.07 (155)	7.11 ± 0.07 (78)	6.94 ± 0.07 (114)
	Diprenorphine	7.88 ± 0.06 (13)	7.63 ± 0.06 (24)	7.51 ± 0.06 (31)	7.46 ± 0.07 (35)
	Etorphine	7.06 ± 0.08 (88)	6.21 ± 0.08 (613)	6.22 ± 0.07 (601)	5.84 ± 0.07 (1569)
	Naltrindole	9.05 ± 0.07 (0.9)	9.21 ± 0.06 (0.6)	8.94 ± 0.05 (1.2)	9.11 ± 0.04 (0.8)
Peptide	DADLE	6.56 ± 0.11 (280)	5.58 ± 0.11 (2600)	5.92 ± 0.07 (1200)	5.23 ± 0.07 (5900)

Displacement of ^3H -naltrindole in the absence and presence (140 mM) of sodium chloride from wild-type δ -OR (WT δ -OR) expressed in HEK293 cells, and the crystallized construct BRIL- δ OR(Δ N/ Δ C)-naltrindole expressed in Sf9 insect cells. Data were analysed using a one-site model. Values represented as pK_i average \pm s.e.m. (K_i nM) from minimum of two assays, each in triplicate.

Extended Data Table 3 | Effects of various mutations on δ -OR radioligand binding parameters**a**

δ -OR	^3H -naltrindole K_d (nM)	DADLE binding affinity against ^3H -naltrindole		
		$[^3\text{H}\text{-naltrindole}]$ nM in assays	$pK_i \pm \text{SEM}$ (n)	Average K_i (μM)
Wild-type	0.52	1.25 – 1.60	6.47 ± 0.09 (4)	0.34
D95A	0.68	1.25 – 1.60	5.51 ± 0.04 (5)	3.09
D95N	0.96	1.25 – 1.60	5.45 ± 0.05 (5)	3.52
D128A	NA	NA	NA	NA
N131A	1.17	1.25 – 1.60	7.12 ± 0.06 (5)	0.08
N131V	1.17	1.25 – 1.60	6.83 ± 0.06 (5)	0.15
S135A	NA	NA	NA	NA
N310A	NA	NA	NA	NA
N314A	NA	NA	NA	NA

b

δ -OR	High affinity fraction	High pK_i (95% C.I.)	Low pK_i (95% C.I.)
Wild-type	0.70 ± 0.05	6.90 (7.06 – 6.73)	5.10 (5.48 – 4.72)
D95A	0.49 ± 0.19	6.07 (6.55 – 5.59)*	4.95 (5.42 – 4.49)
D95N	$0.17 \pm 0.05^\dagger$	6.72 (7.32 – 6.11)	5.14 (5.26 – 5.04)
N131A	0.69 ± 0.12	7.46 (7.59 – 7.34)*	4.57 (4.85 – 4.29)
N131V	0.75 ± 0.03	7.24 (7.35 – 7.13)*	4.81 (5.15 – 4.47)

a. For these experiments ^3H -naltrindole saturation and DADLE competition assays using ^3H -naltrindole in the absence of sodium chloride were performed. Results are presented as average \pm s.e.m. from two or more separate experiments, each in triplicate. Binding curves were fit to a one-site model. The same data were also fit to a two-site model and the results are shown in Extended Data Table 3b. NA, no activity; these receptor mutants expressed receptor which did not bind ^3H -naltrindole. **b.** DADLE binding against ^3H -naltrindole in the absence of sodium chloride fit to a two-site model. Results are presented as average \pm s.e.m. from three or four assays. High-affinity fraction values were taken directly from curve-fitting. High and low IC_{50} values and their corresponding 95% confidence intervals (C.I.) were taken from curve-fitting and converted to corresponding high and low affinity (pK_i) and 95% C.I. values using the Cheng–Prusoff equation. Note that D95A reduced high DADLE-binding affinity, but did not significantly reduce the high-affinity fraction. D95N reduced the high-affinity fraction, but did not reduce DADLE high affinity. N131A or N131V increased DADLE high affinity. DADLE low affinity remained unchanged in all mutants. * $P < 0.0001$, ** $P < 0.05$, as compared with wild-type δ -OR using one-way ANOVA.

The observable signature of late heating of the Universe during cosmic reionization

Anastasia Fialkov^{1,2}, Rennan Barkana¹ & Eli Visbal^{3,4,5}

Models and simulations^{1–4} of the epoch of reionization predict that spectra of the 21-centimetre transition of atomic hydrogen will show a clear fluctuation peak, at a redshift and scale, respectively, that mark the central stage of reionization and the characteristic size of ionized bubbles. This is based on the assumption^{5–7} that the cosmic gas was heated by stellar remnants—particularly X-ray binaries—to temperatures well above the cosmic microwave background at that time (about 30 kelvin). Here we show instead that the hard spectra (that is, spectra with more high-energy photons than low-energy photons) of X-ray binaries^{8,9} make such heating ineffective, resulting in a delayed and spatially uniform heating that modifies the 21-centimetre signature of reionization. Rather than looking for a simple rise and fall of the large-scale fluctuations (peaking at several millikelvin), we must expect a more complex signal also featuring a distinct minimum (at less than a millikelvin) that marks the rise of the cosmic mean gas temperature above the microwave background. Observing this signal, possibly with radio telescopes in operation today, will demonstrate the presence of a cosmic background of hard X-rays at that early time.

Although stellar remnants at high redshift have been considered previously, a more reliable prediction of the radiative feedback from X-ray binaries (XRBs) is now possible, owing to a recent detailed population synthesis simulation of their evolution over cosmic time^{8,9}. This simulation was calibrated to all available observations in the local and low-redshift Universe, and it predicts the evolution of the luminosity and the X-ray spectrum of XRBs with redshift. In particular, high-mass XRBs (especially black-hole binaries) should dominate, with a ratio at high redshift of bolometric X-ray luminosity to the star-formation rate (SFR) of

$$\frac{L_X}{\text{SFR}} = 3 \times 10^{40} f_X \text{ erg s}^{-1} M_{\odot}^{-1} \text{ yr} \quad (1)$$

We have allowed for an uncertainty in the X-ray efficiency with an extra parameter in equation (1), where $f_X = 1$ indicates our standard value. We focus on XRBs as the most natural heating source, because other observed sources should be sub-dominant at high redshift (see Methods).

Previous calculations of X-ray heating^{6,10–13} have assumed power-law spectra that place most of the X-ray energy at the low-energy end, where the mean free path of the soft X-rays is relatively short. This means that most of the emitted X-rays are absorbed soon after they are emitted, before much energy is lost as a result of cosmological effects. The absorbed energy is then enough to heat the gas by the time of reionization to about ten times the temperature of the cosmic microwave background (CMB; see Methods). Thus, it is generally assumed that reionization occurs when $T_{\text{gas}} \gg T_{\text{CMB}}$, a limit referred to as ‘saturated heating’, because the 21-cm intensity then becomes independent of T_{gas} and mainly dependent on ionization and density. A different possibility, in which heating is delayed until reionization, has only previously been considered as a fringe case of having an unusually low ratio of X-ray luminosity to SFR^{10,11}.

However, the average radiation from XRBs is expected to have a much harder spectrum (Fig. 1), whose energy content (per logarithmic frequency

interval) peaks at around 3 keV. Photons above a (roughly redshift-independent) critical energy of about 1 keV have such a long mean free path that by the start of reionization, most of these photons have not yet been absorbed, and the absorbed ones came from distant sources that were effectively dimmed as a result of cosmological redshift effects. This reduces the absorbed energy by about a factor of five, enough to push the moment at which the mean gas temperature equals that of the CMB into the expected redshift range of cosmic reionization. This moment (termed the ‘heating transition’) is a key milestone in 21-cm cosmology.

The immediate effect of the newly predicted late heating is to give the cosmic gas more time to cool adiabatically to well below the CMB temperature, thus producing mean 21-cm absorption that reaches a maximum depth in the range -110 mK to -180 mK at $z \approx 15–19$ (Fig. 2). This may make it easier for experiments to detect the global 21-cm spectrum from before reionization and thus probe the corresponding early galaxies. Global experiments are most sensitive to the frequency

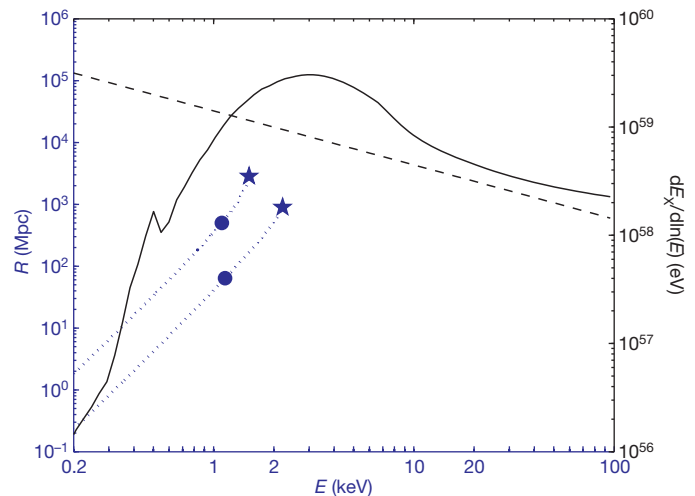


Figure 1 | X-ray spectra, mean free paths, and horizons. We compare the expected spectrum of XRBs at high redshift (solid curve) from population synthesis models^{8,9} to the soft power-law spectrum (dashed line) adopted in previous studies^{6,10–13}. Both indicate the distribution into X-ray photons with energy E of the total X-ray energy E_X produced per solar mass of newly formed stars, for $f_X = 1$ in equation (1). The X-ray emission of XRBs should be dominated by the most massive systems in their high (that is, bright) state⁹, which is dominated by thermal disk emission, with little emission expected or seen^{8,27,28} below about 1 keV. We also show the mean free paths (dotted blue curves) of X-ray photons arriving at $z = 10$ (upper curve) or $z = 30$ (lower curve). For each of these redshifts, we also indicate the effective horizon (defined as a $1/e$ drop-off, like a mean free path) from the combined effect of cosmological redshift and time retardation of sources (blue circle), and the distance to $z = 65$ (blue star), the formation redshift of the first star^{29,30} (where we cut off the mean free path curves). Note the separate y axes, which indicate energy content for the spectra (right) and comoving distance for the other quantities (left).

¹Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. ²Département de Physique, Ecole Normale Supérieure, CNRS, 24 rue Lhomond, 75005 Paris, France. ³Department of Astronomy, Columbia University, 550 West 120th Street, New York, New York 10027, USA. ⁴Jefferson Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

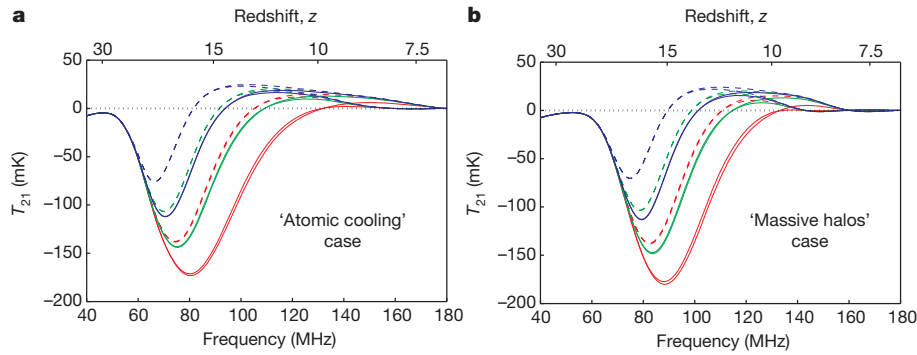


Figure 2 | The global 21-cm spectrum. We show the cosmic mean 21-cm brightness temperature (T_{21}) relative to the CMB versus observed frequency, for the new XRB spectrum⁹ (solid curves) and for the previously adopted soft spectrum (dashed curves); note also the fiducial dotted line at $T_{21} = 0$. We consider our standard $f_X = 1$ case (green curves), as well as $f_X = 1/\sqrt{10}$ (red curves) and $f_X = \sqrt{10}$ (blue curves), each with either early or late reionization (the ‘early’ case is always closer to $T_{21} = 0$). We consider our two cases for galactic halos, atomic cooling (a) or massive halos (b), where the latter has a higher star-formation efficiency (see Methods). With the old spectrum, the prediction with $f_X = 1$ was for a deepest absorption signal of $T_{21} = -107$ mK to -103 mK at $z = 17-19$ (ranges indicate our various halo

derivative of the 21-cm brightness temperature; late heating extends the steep portion of the spectrum to higher frequencies, moving the maximum positive derivative to a frequency about 10% higher (where the foregrounds are significantly weaker) while also changing the value of this maximum derivative by $\pm 10\%$.

We illustrate the possibilities for the timing of reionization with two example cases: our ‘early’ reionization case, which has an optical depth to reionization equal to the most likely value according to CMB experiments^{14,15}, and our ‘late’ reionization case, which is 1σ below the central optical depth value and is more in line with possible hints of a late end to reionization¹⁶. In each case we also consider two possibilities

and reionization cases), and a minimum mean $T_{\text{gas}} = 8.4-10.4$ K at $z = 18.5-20$. With the new spectrum, these values change to $T_{21} = -148$ mK to -143 mK at $z = 16-18$, and $T_{\text{gas}} = 6.9-8.4$ K at $z = 16.5-18$. Star formation in the models is normalized based on reionization, so the ‘massive halos’ case has less star formation and heating (than does the ‘atomic cooling’ case) before reionization, and it produces a sharper global 21-cm signal. These plots extend to higher redshift than is our main focus in this paper, so the high-redshift (that is, low-frequency) drop to the trough could begin at somewhat lower frequencies than indicated if there is a significant contribution from lower-mass halos that we have not included here²⁶.

for the dark matter halos that host galaxies. We focus here on reionization, by which time star formation in the small host halos of the first stars has been shut off by Lyman–Werner radiation^{17–19}, so we consider a minimum mass set by the need for efficient atomic cooling, or a minimum mass that is higher by an order of magnitude; the latter is an example of the case where lower-mass halos are inefficient at star formation, due, for example, to internal supernova feedback^{20,21}. We refer to these two cases as ‘atomic cooling’ and ‘massive halos’, respectively.

The second key consequence of X-ray heating by a hard spectrum is a suppression of 21-cm fluctuations due to heating. Under the previously assumed soft spectra, the short typical distance travelled by the

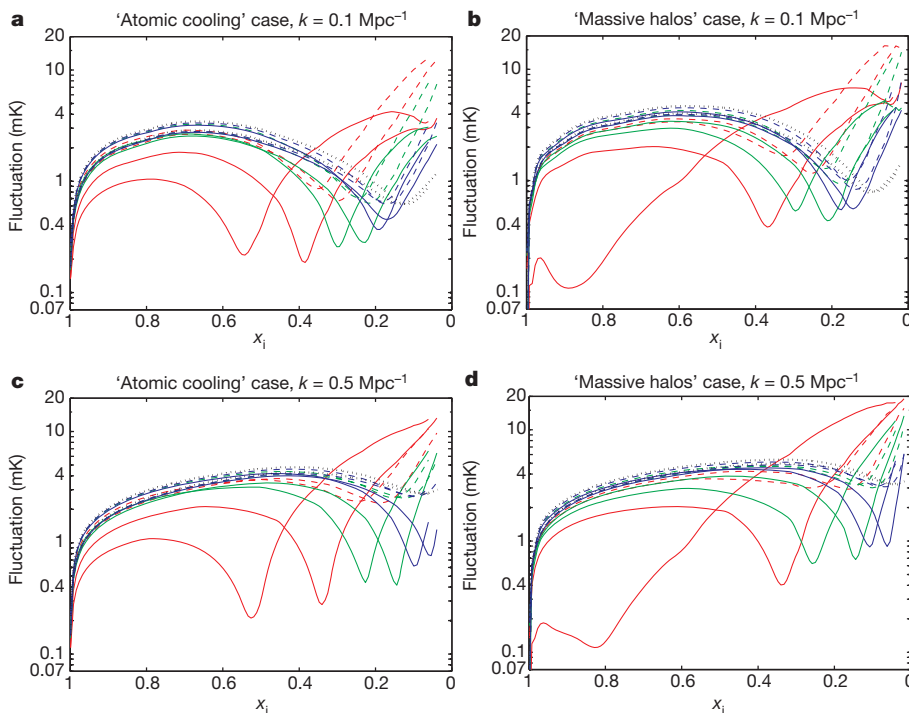


Figure 3 | The 21-cm power spectrum. We show the fluctuation level, defined as $[k^3 P(k)/(2\pi^2)]^{1/2}$ in terms of the power spectrum $P(k)$ of the 21-cm brightness temperature fluctuations, versus the ionized (mass) fraction of the universe x_i (starting on the right from $z = 15$). We compare the new XRB spectrum⁹ (solid curves) to the previously adopted soft spectrum (dashed curves), and show the ‘saturated heating’ case for reference (black dotted curves). We consider our standard $f_X = 1$ case (green curves), as well as $f_X = 1/\sqrt{10}$ (red curves) and $f_X = \sqrt{10}$ (blue curves), each with either early or late reionization. We consider wavenumber $k = 0.1 \text{ Mpc}^{-1}$ (a and b) or $k = 0.5 \text{ Mpc}^{-1}$ (c and d), for each of our two cases for galactic halos, ‘atomic cooling’ (a and c) or ‘massive halos’ (b and d). Our lower k value roughly tracks large-scale fluctuations (heating early on, and ionized bubbles later), whereas our higher k value corresponds to a smaller scale (though one that can still be measured accurately with current experiments) and thus tracks more closely the evolution of density fluctuations. To illustrate the effect of the X-ray spectrum on the results, consider the fluctuation level at $k = 0.5 \text{ Mpc}^{-1}$ at the midpoint of reionization (that is, $x_i = 0.5$); the parameter space we explore gives a possible range of 3.6–4.9 mK for the old spectrum, whereas the new spectrum gives a much broader range of 0.3–4.4 mK. The latter values are typically much lower than the often-assumed limit of saturated heating (which gives a corresponding range of 4.1–5.1 mK).

X-ray photons was found to produce large fluctuations in the gas temperature and thus in the 21-cm intensity around the time of the heating transition, regardless of when this transition occurred^{10,11,22}. However, the larger source distances associated with a hard spectrum lead to a much more uniform heating, with correspondingly low temperature fluctuations even around the time of the heating transition, when the 21-cm intensity is most sensitive to the gas temperature. This trend is strengthened by late heating, because it occurs at a time when the heating sources are no longer as rare and as strongly biased as they would be in the case of an earlier heating era. Thus, heating with a hard X-ray spectrum is predicted to produce a new signature in the 21-cm fluctuation signal: a deep minimum during reionization (Fig. 3; also shown versus redshift in Extended Data Fig. 1). This results from the low level of gas temperature fluctuations in combination with a suppression of the 21-cm impact of other types of fluctuation (that is, in density and ionization); in particular, right at the heating transition, the cosmic mean 21-cm intensity is (very nearly) zero, and thus all fluctuations other than those in gas temperature disappear (to linear order) from the 21-cm sky.

Depending on the parameters, the deep minimum (reaching below 1 mK) may occur at any time during reionization, but is likely to occur before its midpoint. Previously, the fluctuation signal was expected to lie within a narrow, well-defined range, allowing for a straightforward interpretation of the data in terms of the progress of reionization; now, however, there is a variety of possibilities (Fig. 3), so modelling of data will involve an analysis of the interplay of heating and reionization.

Although ongoing experiments hope to reach a sub-millikelvin sensitivity level^{7,23}, the best current upper limit²⁴ of 52 mK at $k = 0.075 \text{ Mpc}^{-1}$ at $z = 7.7$ is two orders of magnitude away from our predictions. If a sufficient sensitivity level can be achieved, a low minimum in the 21-cm power spectrum during reionization would be a clear signature of late heating due to a hard X-ray spectrum. Indeed, a clear observational indication that this feature corresponds to a cosmic milestone is that the minima at all $k > 0.5 \text{ Mpc}^{-1}$ should occur at essentially the same redshift (namely, the true redshift of the heating transition); the minima at lower wavenumbers should be delayed owing to rapid evolution in the power spectrum shape (see Extended Data Fig. 2).

Heating by high-energy X-rays would remove the previously expected signal from an early heating transition^{10,22} at $z \approx 15 - 20$, but would leave in place the similar $z \approx 25$ signal from the Lyman- α coupling transition that is probably detectable with the Square Kilometre Array^{25,26}. It could also affect other observations of high-redshift galaxies. For example, because late heating implies weak photoheating feedback, low-mass halos may continue to produce copious stars in each region right up to its local reionization, although internal feedback (arising from supernovae or mini-quasars) could still limit star formation in small halos.

METHODS SUMMARY

Our calculations are performed with a hybrid, semi-numerical method that produces realistic, three-dimensional realizations of the early galaxies along with self-consistent inhomogeneous distributions of the important radiation backgrounds. This method improves upon our own and others' previous work^{12,19,22,26}. In particular, we have added to our previous code a calculation of reionization (by stellar ultraviolet photons) and of partial ionization by X-rays. Our approach is similar to the 21CMFAST code¹² except that in the heating code we calculate the X-ray optical depth (as a function of position and of frequency) much more accurately.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July; accepted 20 December 2013.

Published online 5 February 2014.

1. Barkana, R. & Loeb, A. Unusually large fluctuations in the statistics of galaxy formation at high redshift. *Astrophys. J.* **609**, 474–481 (2004).
2. Furlanetto, S. R., Zaldarriaga, M. & Hernquist, L. The growth of H II regions during reionization. *Astrophys. J.* **613**, 1–15 (2004).

3. Mellema, G., Iliiev, I. T., Pen, U.-L. & Shapiro, P. R. Simulating cosmic reionization at large scales—II. The 21-cm emission features and statistical signals. *Mon. Not. R. Astron. Soc.* **372**, 679–692 (2006).
4. Zahn, O. *et al.* Simulations and analytic calculations of bubble growth during hydrogen reionization. *Astrophys. J.* **654**, 12–26 (2007).
5. Madau, P., Meiksin, A. & Rees, M. J. 21 centimeter tomography of the intergalactic medium at high redshift. *Astrophys. J.* **475**, 429–444 (1997).
6. Furlanetto, S. R. The global 21-centimeter background from high redshifts. *Mon. Not. R. Astron. Soc.* **371**, 867–878 (2006).
7. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. Cosmology at low frequencies: the 21 cm transition and the high-redshift Universe. *Phys. Rep.* **433**, 181–301 (2006).
8. Fragos, T. *et al.* X-ray binary evolution across cosmic time. *Astrophys. J.* **764**, 41 (2013).
9. Fragos, T. *et al.* Energy feedback from X-ray binaries in the early Universe. *Astrophys. J.* **776**, L31 (2013).
10. Pritchard, J. R. & Furlanetto, S. 21-cm fluctuations from inhomogeneous X-ray heating before reionization. *Mon. Not. R. Astron. Soc.* **376**, 1680–1694 (2007).
11. Christian, P. & Loeb, A. Measuring the X-ray background in the reionization era with first generation 21 cm experiments. *J. Cosmol. Astroparticle Phys.* **09**, 014 (2013).
12. Mesinger, A., Furlanetto, S. & Cen, R. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *Mon. Not. R. Astron. Soc.* **411**, 955–972 (2011).
13. Mesinger, A., Ferrara, A. & Spiegel, D. S. Signatures of X-rays in the early Universe. *Mon. Not. R. Astron. Soc.* **431**, 621–637 (2013).
14. Bennett, C. L. *et al.* Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: final maps and results. *Astrophys. J.* **208** (suppl.), 20 (2013).
15. Planck Collaboration. Planck 2013 results. XVI. Cosmological parameters. Preprint at <http://arxiv.org/abs/1303.5076> (2013).
16. Schroeder, J., Mesinger, A. & Haiman, Z. Evidence of Gunn-Peterson damping wings in high- z quasar spectra: strengthening the case for incomplete reionization at $z \sim 6 - 7$. *Mon. Not. R. Astron. Soc.* **428**, 3058–3071 (2013).
17. Haiman, Z., Rees, M. J. & Loeb, A. Destruction of molecular hydrogen during cosmological reionization. *Astrophys. J.* **476**, 458–463 (1997); erratum *Astrophys. J.* **484**, 985 (1997).
18. Ahn, K. *et al.* Detecting the rise and fall of the first stars by their impact on cosmic reionization. *Astrophys. J.* **756**, L16 (2012).
19. Fialkov, A., Barkana, R., Visbal, E., Tselikhovich, D. & Hirata, C. M. The 21-cm signature of the first stars during the Lyman-Werner feedback era. *Mon. Not. R. Astron. Soc.* **432**, 2909–2916 (2013).
20. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303**, 39–55 (1986).
21. Wyithe, J. S. B. & Loeb, A. A suppressed contribution of low-mass galaxies to reionization due to supernova feedback. *Mon. Not. R. Astron. Soc.* **428**, 2741–2754 (2013).
22. Visbal, E., Barkana, R., Fialkov, A., Tselikhovich, D. & Hirata, C. M. The signature of the first stars in atomic hydrogen at redshift 20. *Nature* **487**, 70 (2012).
23. McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L. & Furlanetto, S. R. Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *Astrophys. J.* **653**, 815–834 (2006).
24. Parsons, A. R. *et al.* New limits on 21cm EoR from PAPER-32 consistent with an X-ray heated IGM at $z = 7.7$. Preprint at <http://arxiv.org/abs/1304.4991> (2013).
25. Barkana, R. & Loeb, A. Detecting the earliest galaxies through two new sources of 21 centimeter fluctuations. *Astrophys. J.* **626**, 1–11 (2005).
26. Fialkov, A., Barkana, R., Pinhas, A. & Visbal, E. Complete history of the observable 21-cm signal from the first stars during the pre-reionization era. *Mon. Not. R. Astron. Soc.* **437**, L36–L40 (2014).
27. McClintock, J. E. & Remillard, R. A. in *Compact Stellar X-Ray Sources* (eds Lewin, W. & van der Klis, M.) 157–213 (Cambridge Astrophys. Ser. 157, Cambridge Univ. Press, 2006).
28. Tamura, M. *et al.* The truncated disk from Suzaku data of GX 339–4 in the extreme very high state. *Astrophys. J.* **753**, 65 (2012).
29. Naoz, S., Noter, S. & Barkana, R. The first stars in the Universe. *Mon. Not. R. Astron. Soc.* **373**, L98–L102 (2006).
30. Fialkov, A., Barkana, R., Tselikhovich, D. & Hirata, C. Impact of the relative motion between dark matter and baryons on the first stars: semi-analytical modelling. *Mon. Not. R. Astron. Soc.* **424**, 1335–1345 (2012).

Acknowledgements We are grateful to S. Naoz for drawing our attention to the work of T. Fragos, who provided us with detailed model spectra of X-ray binaries, which helped motivate this study. This work was supported by Israel Science Foundation grant number 823/09, and by the LabEx ENS-ICFP (grant numbers ANR-10-LABX-0010 and ANR-10-IDEX-0001-02 PSL*).

Author Contributions R.B. initiated the project. A.F. developed and ran the simulations and made the figures by substantially extending a code that was originally developed by E.V. working with R.B. The text was written by R.B. and edited by the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.F. (anastasia.fialkov@gmail.com) or R.B. (barkana@wise.tau.ac.il).

METHODS

Hybrid simulation code. For this work we extended the independent code that we had previously developed^{19,22,26}. This code implements a hybrid, semi-numerical method to produce instances of the expected three-dimensional distribution of early star-forming halos³¹. We first used the known statistical properties of the initial density perturbations to generate a realistic sample universe on large, linear scales. Specifically, we assumed Gaussian initial conditions and adopted the initial power spectrum corresponding to the currently best-measured cosmological parameters from the Planck satellite¹⁵. In a cubic volume consisting of 128^3 cells (each three comoving Mpc on a side), we generated, as in our previous work, a random realization of the initial overdensity (with periodic boundary conditions). Our code also fully includes the effect of the spatially varying relative baryon–dark matter velocity^{22,32} as well as inhomogeneous Lyman–Werner feedback^{17–19}, but these play a very minor part in this paper, because here we focused on relatively late times at which star formation is dominated by relatively high-mass halos.

Given the large-scale density distribution, we then computed the gas fraction in star-forming halos in each cell as a function of time, as in our previous papers. Specifically, this gas has density

$$\rho_{\text{gas}} = \int_{M_{\text{cool}}}^{\infty} \frac{dn}{dM} M_{\text{gas}}(M) dM \quad (2)$$

where dn/dM is the comoving abundance of halos of mass M (that is, n is the comoving number density), $M_{\text{gas}}(M)$ is the gas mass inside a halo of total mass M , and M_{cool} is the minimum halo mass in which the gas can cool efficiently and form stars. The stellar density equals ρ_{gas} multiplied by the star-formation efficiency.

We then used this information to determine the X-ray heating rate in each cell as follows. At each redshift, we smoothed the stellar density field in shells around each cell, by filtering it (using fast Fourier transforms) with two position-space top-hat filters of different radii and taking the difference. We assumed the flux of X-ray photons emitted from each shell to be proportional to the SFR (equation (1)), which is in turn proportional to the time derivative of ρ_{gas} . We then computed the heating by integrating over all the shells seen by each cell, as in the 21CMFAST code¹². In this integral, the radiative contribution of each cell to a given central cell was computed at the time-delayed redshift seen by the central cell, using a pre-computed interpolation grid of star formation versus overdensity and redshift. We varied the number and thickness of shells to check for convergence. We used photo-ionization cross-sections and energy deposition fractions from atomic physics calculations^{33,34}.

Given the X-ray heating rate versus redshift at each cell, we integrated as in 21CMFAST to get the gas temperature as a function of time. We interpolated the heating rate between the redshifts at which it was explicitly computed, and varied the number of redshifts to ensure convergence. Our code also fully includes the inhomogeneous coupling between the 21-cm spin temperature and the gas temperature due to Lyman- α radiation²⁶, but in this paper we focused on late times when the Lyman- α coupling has nearly saturated, and thus the 21-cm spin temperature equals the kinetic gas temperature to high accuracy. However, the incomplete saturation has a small effect at the high-redshift ($z \approx 15$) end of Fig. 3 and Extended Data Figs 1 and 2; also, it is the main driver of the high-redshift portion of the global signal shown in Fig. 2.

The 21-cm brightness temperature (relative to the CMB temperature T_{CMB}) equals⁵

$$T_{21} = 29(1 + \delta)x_{\text{H I}} \left(1 - \frac{T_{\text{CMB}}}{T_{\text{S}}}\right) \sqrt{\frac{1+z}{10}} \text{ mK} \quad (3)$$

where T_{S} is the spin temperature of hydrogen, δ is the overdensity, and $x_{\text{H I}}$ is the neutral fraction of hydrogen. In our code we also included a small modification of equation (3) at low gas temperatures^{35,36}. We calculated reionization maps as in 21CMFAST¹², accounting for the effect of large-scale galaxy fluctuations¹ by setting each cell to be fully reionized if some sphere around it contains enough ionizing photons to self-reionize²; we considered spheres up to 70 comoving Mpc, roughly the maximum size of ionized bubbles at the end of reionization³⁷. We took an ionization threshold corresponding to three ionizations of each atom due to recombinations. This number can be lowered without changing our results if the escape fraction of ionizing photons is correspondingly lowered as well. In any case, as mentioned in the main text, we set the effective ionization efficiency (in our early and late reionization cases) according to empirical limits on reionization.

As also noted in the main text, we considered two possibilities for the minimum mass of dark matter halos that host galaxies. Specifically, we considered a minimum mass set by the need for efficient atomic cooling (corresponding to a minimum halo circular velocity of 16.5 km s^{-1}), or a massive halo case with a minimum mass higher by a factor of 9.5 (corresponding to a minimum circular velocity of

35 km s^{-1}). We adopted a star formation efficiency of $f_* = 0.05$ for our atomic cooling case and $f_* = 0.15$ for massive halos, so that in our model both cases require an escape fraction of ionizing photons of approximately 20% for our late reionization case and approximately 40% for early reionization. Taking lower f_* would push the escape fraction towards unrealistically high values, whereas higher f_* would strongly contradict recent numerical simulations that suggest relatively low values for such halos^{38,39}.

Other than reionization, another addition to our code (compared to our previous publications) was partial ionization by the same X-rays that heat the gas. These X-rays make a negligible contribution to reionization compared to stellar ultraviolet photons, but their slight ionization of gas that is still mostly neutral affects heating through the ionization dependence of various cross-sections³³. In the calculation of heating and ionization by X-rays, our code accurately follows the inhomogeneous X-ray optical depth (though with the approximation of a cosmic mean density) between sources and absorbers. In particular, it includes two substantial advantages over 21CMFAST¹². First, when the simulated universe begins to reionize, we do not use the cosmic mean ionization fraction for the optical depth, but instead use the local, spherically averaged value (including time-retardation) in the region between each spherical shell of X-ray sources and the central, absorbing pixel. This means that our hybrid simulation correctly heats the gas immediately surrounding H II regions especially strongly, owing both to the presence of an overdensity of sources nearby (within the ionized regions) and the low optical depth to these sources (because most of the intervening gas is ionized). And, second, we do not use the two-step approximation of optical depth in 21CMFAST, whereby the effective optical depth is either 0 or ∞ , but instead account for the correct (spherically averaged) optical depth seen by photons of each frequency from each spherical shell.

Rough estimate of X-ray heating. We can understand the critical parameters involved in cosmic heating with a simple estimate. The amount of energy inserted into the gas by X-ray sources is determined by the collapse fraction f_{coll} (that is, the fraction of the cosmic gas that makes it into star-forming dark-matter halos), the star-formation efficiency f_* , the fraction f_{abs} of the X-ray energy that is absorbed by the gas (after losses due to redshift effects), and the fraction f_{heat} of the absorbed X-ray energy that goes into heating. A simple estimate of the effect of X-ray heating is given by equating the X-ray heating energy with the thermal energy of the gas, yielding the resulting gas temperature (neglecting adiabatic and Compton cooling or heating). At redshift z , the ratio of this gas temperature from X-ray heating to the CMB temperature is:

$$\frac{T_{\text{X}}}{T_{\text{CMB}}} \approx 9f_{\text{X}} \frac{f_{\text{coll}}}{0.01} \times \frac{f_*}{0.05} \times \frac{f_{\text{abs}}}{0.8} \times \frac{f_{\text{heat}}}{0.12} \times \frac{10}{1+z} \quad (4)$$

where we have assumed the L_{X}/SFR ratio from equation (1). Thus, it is generally assumed that reionization occurs when $T_{\text{gas}} \gg T_{\text{CMB}}$, that is, much later than the heating transition (when the mean gas temperature equals that of the CMB), in the regime of saturated heating when the 21-cm emission is independent of the gas temperature. As explained in the main text, the new XRB spectrum that we adopt⁹ reduces f_{abs} by a factor of five and changes this conclusion.

We note that this estimate assumes our standard ratio in equation (1), which includes an order-of-magnitude increase in this ratio at the low metallicity expected for high-redshift galaxies compared to the local ratio at solar metallicity. This increase is predicted by the evolutionary models and suggested (though not yet solidly confirmed) by observations^{9,40–42}; not including it would only strengthen our case for a late heating. We also note that our standard ratio includes intrinsic (that is, interstellar) absorption, assuming that the strength of this absorption at high redshift is similar to that at low redshift⁹.

A similar value for the L_{X}/SFR ratio has been claimed previously for local galaxies⁶; this high value was reached by extrapolating hard (2–10 keV) X-ray observations down to 0.2 keV according to the previously adopted soft spectrum (Fig. 1), resulting in an overestimate by a factor of about three. This leaves a further factor-of-three discrepancy (which may mostly be due to the difference between the SFR of high-mass stars, as sometimes used⁴³, compared to the total SFR). We also note that we have used in equation (4) the rather low value of f_{heat} appropriate for gas³³ with the residual ionized fraction of around 2×10^{-4} expected before X-rays begin to raise this value.

Other heating sources. In this paper we have focused on X-ray binaries as the most natural high-redshift heating source, because they are a direct consequence of the star formation believed to be responsible for reionizing the Universe, and high-redshift XRB populations should be generally similar to those observed in the local Universe (except for some differences due to metallicity). We note that XRBs from the high redshifts considered here make only a small contribution to the observed X-ray background⁹.

On the basis of low-redshift observations, other potential X-ray sources appear sub-dominant compared to XRBs. One such source is thermal emission from hot

gas in galaxies. The ratio of its X-ray luminosity to the SFR in local galaxies⁴⁴ is about 10^{39} (including intrinsic absorption). This is well below even the local ratio for XRBs, so it should make only a minor contribution at high redshift (compared with equation (1)). Some theoretical arguments suggest that X-rays produced via Compton emission from relativistic electrons could be important at high redshift⁴⁵, but the expected spectrum (flat from approximately 100 eV to 100 GeV) would deliver most of its energy above 1 keV and thus produce 21-cm signatures similar to those of our assumed XRB spectrum.

Another possible X-ray source is the population of bright quasars. Although quasars are believed to dominate the X-ray background at low redshift⁴⁶, their rapid decline beyond $z \approx 3$ suggests that their total X-ray luminosity (including an extrapolation of their observed luminosity function) is sub-dominant compared to XRBs during and before reionization⁹.

More promising, perhaps, is the possibility of a population of mini-quasars—that is, central black holes in early star-forming halos. This must be considered speculative, because the early halos are so small compared to galactic halos in the present Universe that the corresponding black-hole masses should fall in a very different range from observed quasars, specifically within the intermediate black-hole range ($10^2 - 10^4 M_\odot$) that local observations have probed only to a limited extent⁴⁷. Thus, the properties of these mini-quasars are highly uncertain, and various assumptions can allow them to produce either early or late heating^{48,49}.

We can use local observations to try to estimate the possible importance of mini-quasars. An internal feedback model that is consistent with observations of local black-hole masses as well as high-redshift quasar luminosity functions suggests the following relation between the black-hole mass M_{bh} and the mass of its host dark-matter halo M_{halo} (ref. 50):

$$M_{\text{bh}} = 1.5 \times 10^{-6} M_{\text{halo}} \left(\frac{M_{\text{halo}}}{10^8 M_\odot} \right)^{2/3} \left(\frac{1+z}{10} \right)^{5/2} \quad (5)$$

Assuming that mini-quasars shine at the Eddington luminosity, and that this luminosity comes out in X-rays, gives

$$L_X = 1.5 \times 10^{40} \text{ erg s}^{-1} \times \left(\frac{M_{\text{halo}}}{10^8 M_\odot} \right)^{5/3} \left(\frac{1+z}{10} \right)^{5/2} \quad (6)$$

Next, we assume that each mini-quasar shines at this luminosity for a lifetime equal to the dynamical time of its galactic disk and that the time between active periods is given by the merger timescale⁵⁰. Assuming that this merger timescale is roughly equal to the star-formation timescale (which is defined as the ratio between the total mass in stars to the SFR), we find the ratio of the average X-ray luminosity of the central mini-quasar to that of XRBs in the same halo to be

$$\frac{L_X^{\text{mini-quasar}}}{L_X^{\text{XRBs}}} = 0.2 \left(\frac{f_X f_*}{0.05} \right)^{-1} \left(\frac{M_{\text{halo}}}{10^8 M_\odot} \right)^{2/3} \left(\frac{1+z}{10} \right) \quad (7)$$

where we assumed that on average a quarter of the mini-quasar X-rays in the relevant wavelengths make it past interstellar absorption, similar to observed quasars⁵¹ and XRBs^{9,44}. Thus, this observationally based estimate indicates a small mini-quasar contribution for our atomic cooling case as well as for massive halos (in the massive halo case, the increased ratio from the larger typical halo mass is mostly cancelled out by the increased star-formation efficiency needed).

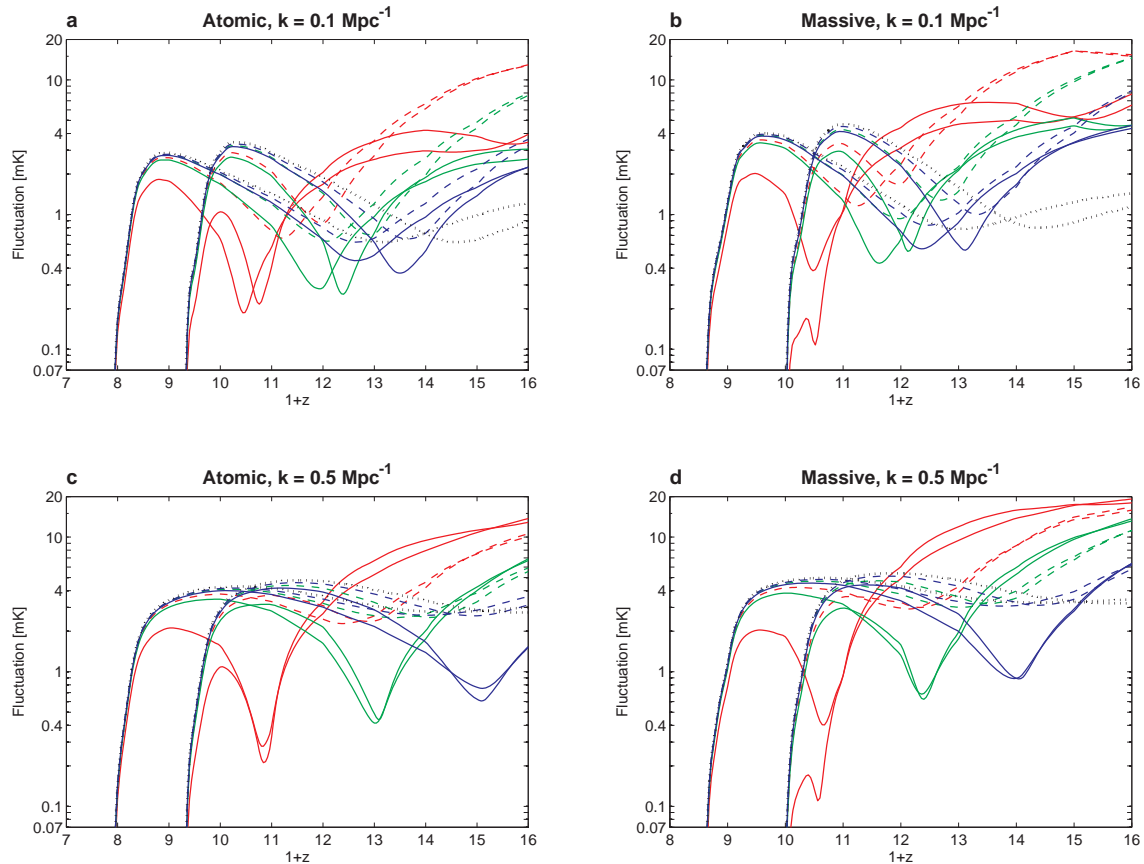
We note that a shallower relation between M_{bh} and the halo velocity dispersion than that assumed in equation (5) would increase the ratio in equation (7) by about an order of magnitude⁵²; however, current data strongly favours a steep relation⁵³. We also note that standard models of accretion disks⁵⁴ around black holes predict that the X-ray spectrum of mini-quasars⁴⁸ should peak at 1–5 keV, making it a hard spectrum that is much closer to the XRB spectrum we have adopted in this paper than to the soft spectrum usually assumed in previous calculations of X-ray heating. Thus, an unusually large contribution from mini-quasars could produce a somewhat earlier heating transition but it would most probably still be marked by a clear minimum in the 21-cm fluctuations.

Regardless of the source of X-rays, an important parameter is the degree of absorption in high-redshift halos compared to locally observed galaxies. If we assume that the gas density in high-redshift halos increases proportionally with the cosmic mean density, then the column density through gas (within a galaxy or a halo) is proportional to $(1+z)^2 M_{\text{halo}}^{1/3}$. This simple relation suggests that absorption of X-rays should increase at high redshift, because the redshift dependence should

have a stronger effect than the decrease of the typical halo mass. However, complex astrophysics could substantially affect this conclusion, because the lower binding energy of the gas in low-mass halos could make it easier to clear out more of the blockading gas. In this work we have assumed that the strength of the intrinsic absorption of X-rays from XRBs at high redshift is the same as that observed at low redshift.

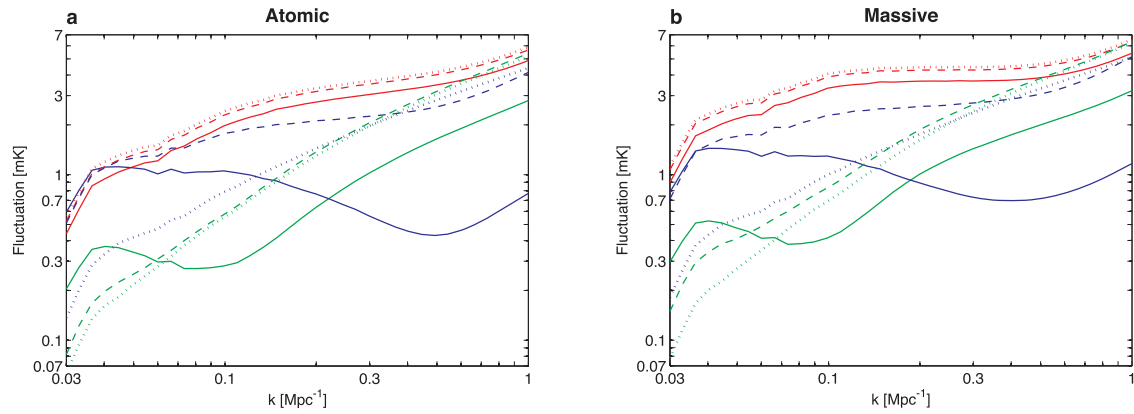
There has also been some speculation about heating sources other than X-ray radiation, but none have been shown to be important. In particular, heating due to Lyman- α photons is very inefficient^{49,55,56}, even compared to the reduced X-ray heating that we find, and shocks can probably heat only a small fraction of the cosmic gas before reionization⁵⁷.

31. Barkana, R. & Loeb, A. In the beginning: the first sources of light and the reionization of the Universe. *Phys. Rep.* **349**, 125–238 (2001).
32. Tselikhovich, D. & Hirata, C. Relative velocity of dark matter and baryonic fluids and the formation of the first structures. *Phys. Rev. D* **82**, 083520 (2010).
33. Furlanetto, S. R. & Stoever, S. J. Secondary ionization and heating by fast electrons. *Mon. Not. R. Astron. Soc.* **404**, 1869 (2010).
34. Verner, D. A., Ferland, G. J., Korista, K. T. & Yakovlev, D. G. Atomic data for astrophysics. II. New analytic FITS for photoionization cross sections of atoms and ions. *Astrophys. J.* **465**, 487 (1996).
35. Hirata, C. M. Wouthuysen-Field coupling strength and application to high-redshift 21-cm radiation. *Mon. Not. R. Astron. Soc.* **367**, 259–274 (2006).
36. Chuzhoy, L. & Shapiro, P. R. Ultraviolet pumping of hyperfine transitions in the light elements, with application to 21 cm hydrogen and 92 cm deuterium lines from the early Universe. *Astrophys. J.* **651**, 1 (2006).
37. Wyithe, J. S. B. & Loeb, A. A characteristic size of ~ 10 Mpc for the ionized bubbles at the end of cosmic reionization. *Nature* **432**, 194 (2004).
38. Safranek-Shrader, C., Milosavljevic, M. & Bromm, V. Star formation in the first galaxies—II: Clustered star formation and the influence of metal line cooling. Preprint at <http://arxiv.org/abs/1307.1982> (2013).
39. Wise, J. H., Abel, T., Turk, M. J., Norman, M. L. & Smith, B. D. The birth of a galaxy—II. The role of radiation pressure. *Mon. Not. R. Astron. Soc.* **427**, 311–326 (2012).
40. Mirabel, I. F., Dijkstra, M., Laurent, P., Loeb, A. & Pritchard, J. R. Stellar black holes at the dawn of the universe. *Astron. Astrophys.* **528**, A149 (2011).
41. Basu-Zych, A. R. *et al.* The X-ray star formation story as told by Lyman break galaxies in the 4 Ms CDF-S. *Astrophys. J.* **762**, 45 (2013).
42. Basu-Zych, A. R. *et al.* Evidence for elevated X-ray emission in local Lyman break galaxy analogs. *Astrophys. J.* **774**, 152 (2013).
43. Gilfanov, M., Grimm, H.-J. & Sunyaev, R. L_X -SFR relation in star-forming galaxies. *Mon. Not. R. Astron. Soc.* **347**, L57 (2004).
44. Mineo, S., Gilfanov, M. & Sunyaev, R. X-ray emission from star-forming galaxies—II. Hot interstellar medium. *Mon. Not. R. Astron. Soc.* **426**, 1870–1883 (2012).
45. Oh, S. P. Reionization by hard photons. I. X-rays from the first star clusters. *Astrophys. J.* **553**, 499 (2001).
46. Vasudevan, R. V., Mushotzky, R. F. & Gandhi, P. Can we reproduce the X-ray background spectral shape using local active galactic nuclei? *Astrophys. J.* **770**, L37 (2013).
47. Lützgendorf, N. *et al.* $M_\bullet - \sigma$ relation for intermediate mass black holes in globular clusters. *Astron. Astrophys.* **555**, A26 (2013).
48. Tanaka, T., Perna, R. & Haiman, Z. X-ray emission from high-redshift miniquasars: self-regulating the population of massive black holes through global warming. *Mon. Not. R. Astron. Soc.* **425**, 2974–2987 (2012).
49. Ciardi, B., Salvaterra, R. & Di Matteo, T. Ly α versus X-ray heating in the high- z intergalactic medium. *Mon. Not. R. Astron. Soc.* **401**, 2635–2640 (2010).
50. Wyithe, J. S. B. & Loeb, A. Self-regulated growth of supermassive black holes in galaxies as the origin of the optical and X-ray luminosity functions of quasars. *Astrophys. J.* **595**, 614 (2003).
51. Sazonov, S., Yu., Ostriker, J. P. & Sunyaev, R. A. Quasars: the characteristic spectrum and the induced radiative heating. *Mon. Not. R. Astron. Soc.* **347**, 144–156 (2004).
52. Volonteri, M. & Gnedin, N. Y. Relative role of stars and quasars in cosmic reionization. *Astrophys. J.* **703**, 2113–2117 (2009).
53. McConnell, N. J. & Ma, C.-P. Revisiting the scaling relations of black hole masses and host galaxy properties. *Astrophys. J.* **764**, 184 (2013).
54. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
55. Chen, X. & Miralda-Escudé, J. The spin-kinetic temperature coupling and the heating rate due to Ly α scattering before reionization: predictions for 21 centimeter emission and absorption. *Astrophys. J.* **602**, 1–11 (2004).
56. Chuzhoy, L. & Shapiro, P. R. Heating and cooling of the early intergalactic medium by resonance photons. *Astrophys. J.* **655**, 843–846 (2007).
57. Furlanetto, S. R. & Loeb, A. Large-scale structure shocks at low and high redshifts. *Astrophys. J.* **611**, 642–654 (2004).



Extended Data Figure 1 | The 21-cm power spectrum versus redshift. We show the same data as in Fig. 3 (with the same nomenclature), but as a function of $1+z$, where z is the redshift, a direct observable (given that the observed wavelength is $21(1+z)$ cm). This presentation has the advantage of clearly separating out the early and late reionization cases, while showing that reionization does not affect the redshift of the new minimum (solid curves) at $k = 0.5 \text{ Mpc}^{-1}$. Indeed, this minimum marks the cosmic heating transition (to within 2% in redshift in all our model calculations), while the minimum at $k = 0.1 \text{ Mpc}^{-1}$ is typically delayed owing to the evolving power spectrum

shape (see Extended Data Fig. 2). We consider wavenumbers $k = 0.1 \text{ Mpc}^{-1}$ (a and b) and $k = 0.5 \text{ Mpc}^{-1}$ (c and d), for each of our two cases for galactic halos, atomic cooling (a and c) or massive halos (b and d). The results shown (here and in Fig. 3) correspond to a total of four different reionization histories. Late reionization with atomic cooling reaches 1/4, 1/2, 3/4 and full reionization at $z = 10.7, 8.7, 7.7$ and 7.0 ; the corresponding redshifts for early reionization are $11.8, 10.0, 9.1$ and 8.4 . Massive halos give a sharper reionization transition, with late reionization advancing through $z = 10.3, 8.9, 8.3$ and 7.7 , while early reionization corresponds to $z = 11.4, 10.2, 9.7$ and 9.0 .



Extended Data Figure 2 | Full 21-cm power spectra. We show examples of full power spectra corresponding to the data shown in Fig. 3 and Extended Data Fig. 1, for the cases with $f_X = 1$ and late reionization. We compare the new XRB spectrum⁹ (solid curves) to the previously adopted soft spectrum (dashed curves), and show the saturated heating case for reference (dotted curves). We consider atomic cooling (a) or massive halos (b). In order of increasing cosmic age, we consider three key moments (which fall at different redshifts for the various cases, based on Extended Data Fig. 1): the minimum

fluctuation at $k = 0.5 \text{ Mpc}^{-1}$ (blue curves), the minimum at $k = 0.1 \text{ Mpc}^{-1}$ (green curves), and the midpoint of reionization (red curves). For the new spectrum, strong evolution is predicted in the power spectrum shape, because large-scale fluctuations from X-ray heating dominate (up to $k \approx 0.5 \text{ Mpc}^{-1}$) at the heating transition (blue solid curves) but then rapidly decline so that density fluctuations come to dominate (at $k > 0.1 \text{ Mpc}^{-1}$ at the time shown in green solid curves), with an eventual large-scale boost by the ionized bubbles (red solid curves).

Hysteresis in a quantized superfluid ‘atomtronic’ circuit

Stephen Eckel¹, Jeffrey G. Lee¹, Fred Jendrzejewski¹, Noel Murray², Charles W. Clark¹, Christopher J. Lobb¹, William D. Phillips¹, Mark Edwards² & Gretchen K. Campbell¹

Atomtronics^{1,2} is an emerging interdisciplinary field that seeks to develop new functional methods by creating devices and circuits where ultracold atoms, often superfluids, have a role analogous to that of electrons in electronics. Hysteresis is widely used in electronic circuits—it is routinely observed in superconducting circuits³ and is essential in radio-frequency superconducting quantum interference devices⁴. Furthermore, it is as fundamental to superfluidity⁵ (and superconductivity) as quantized persistent currents^{6–8}, critical velocity^{9–14} and Josephson effects^{15,16}. Nevertheless, despite multiple theoretical predictions^{5,17–19}, hysteresis has not been previously observed in any superfluid, atomic-gas Bose–Einstein condensate. Here we directly detect hysteresis between quantized circulation states in an atomtronic circuit formed from a ring of superfluid Bose–Einstein condensate obstructed by a rotating weak link (a region of low atomic density). This contrasts with previous experiments on superfluid liquid helium where hysteresis was observed directly in systems in which the quantization of flow could not be observed²⁰, and indirectly in systems that showed quantized flow^{21,22}. Our techniques allow us to tune the size of the hysteresis loop and to consider the fundamental excitations that accompany hysteresis. The results suggest that the relevant excitations involved in hysteresis are vortices, and indicate that dissipation has an important role in the dynamics. Controlled hysteresis in atomtronic circuits may prove to be a crucial feature for the development of practical devices, just as it has in electronic circuits such as memories, digital noise filters (for example Schmitt triggers) and magnetometers (for example superconducting quantum interference devices).

Hysteresis is a general feature of systems where the energy has two (or more) local minima separated by an energy barrier. A schematic of this type of energy landscape is shown in Fig. 1a. A canonical example of hysteresis is the Landau theory of ferromagnetism²³, where the order parameter, κ , is the magnetization, and the energy, $E(\kappa)$, has two minima (stable states) corresponding to the magnetization being aligned parallel or, respectively, antiparallel to the applied magnetic field. For a Bose–Einstein condensate (BEC) in a ring-shaped trap, these minima represent stable flow states of the system, and their energies depend on the applied rotation rate of the trap, Ω (here this rotation is created using a rotating repulsive perturbation). With no interatomic interactions, there is only one minimum in the energy landscape of the BEC. With the addition of interactions, an energy barrier can appear, creating two (or more) stable flow states. This barrier stabilizes the flow, making the BEC a superfluid^{5,24}.

The energy of the barrier is not generally known for superfluid systems; depending on the parameters of the system, it could be related to the energy required to create elementary excitations such as phonons, solitons or vortices. However, the stable states are well known. Rotation of a superfluid in a ring is characterized by a quantized rotation frequency, $n\Omega_0$, where n is the winding number, $\Omega_0 = \hbar/mR^2$ is the rotational quantum, \hbar is Planck’s constant divided by 2π , m is the mass of an atom and R is the mean radius of the trap. The energy of the

superfluid in the frame that rotates with the trap depends on the relative velocity between the superfluid and the trap^{5,24}, and the energy is proportional to $(n - \Omega/\Omega_0)^2$.

Any ring-shaped superfluid necessarily exhibits both hysteresis and a critical rotation rate, Ω_c^\pm (or, equivalently, a critical velocity), because all these effects fundamentally arise from the energy barrier that creates superfluidity. To understand this, we plot the energy of the stable states and the energy barrier as a function of Ω (Fig. 1b, c). Figure 1b shows the ‘swallowtail’ energy structure characteristic of the superfluid. If the system begins in $n = 0$, the flow is stable until $\Omega = \Omega_c^+$, where the energies of the $n = 0$ state and the barrier are equal. At this point, the $n = 0$ state is no longer stable and a transition occurs to $n = 1$, which has lower energy. If Ω is now decreased, this state is stable until $\Omega < \Omega_c^-$, where the flow changes. Thus, a typical hysteresis loop is traced, as shown in the lower panel in Fig. 1b. We note that, although Ω_c^+ and Ω_c^- are the same relative to the superfluid flow, they are generally different in the lab frame and thus appear as hysteresis. Furthermore, in the hysteretic case, they are different from a more general definition of critical rotation (or velocity) that involves the onset

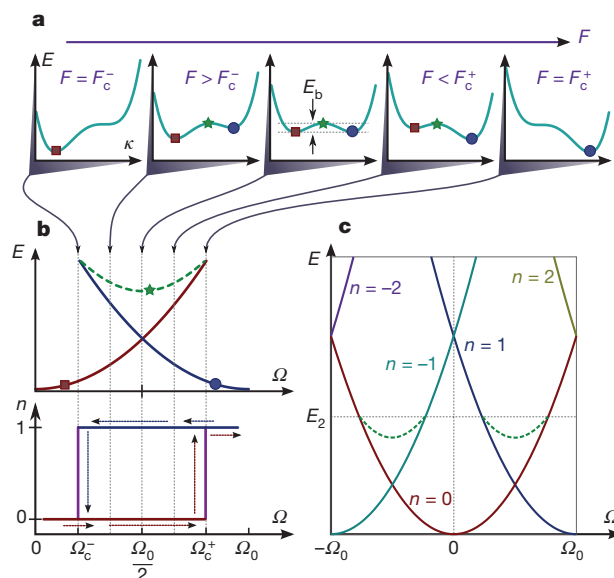


Figure 1 | Origin of hysteresis. **a**, Schematic energy landscape of a hysteretic system. As a function of an order parameter, κ , the energy can have local minima (squares, circles), which represent stable states, separated by a local maximum (stars), which forms an energy barrier, E_b . This landscape is shown for five values of the applied field, F (for superfluidity, $F \rightarrow \Omega$, the rotation rate of the trap). **b**, Plotted as a function of Ω for a superfluid, the energy of the minima (solid) and maximum (dashed) form a swallowtail (top), which exhibits hysteresis (bottom). **c**, This swallowtail structure is periodic in Ω_0 ; states above E_2 are unstable.

¹Joint Quantum Institute, National Institute of Standards and Technology and University of Maryland, Gaithersburg, Maryland 20899, USA. ²Department of Physics, Georgia Southern University, Statesboro, Georgia 30460-8031, USA.

of dissipation or the creation of excitations. At Ω_c^\pm , the hysteretic system may create excitations or experience dissipation, but both cease after the transition is made. Measurement of a hysteresis loop, in addition to measuring Ω_c^\pm , shows an important feature of the underlying energy landscape: the system has at least two stable states. Bi-stability of a moving BEC has been demonstrated independently of quantized states or critical velocities^{25,26}. Finally, unlike that for ferromagnetism, this energy structure is periodic in Ω , with period Ω_0 (Fig. 1c). Similar periodic swallowtail energy structures are predicted for superfluids trapped in a lattice¹⁹.

Our superfluid system is a BEC of ^{23}Na atoms in a ring-shaped optical dipole trap (Fig. 2a). To induce flow, we use a blue-detuned laser to create a rotating repulsive potential, depleting the density in a small portion of the ring and thereby creating a weak link²⁷. The intensity of the laser sets the height of this potential, U . Without this weak link, superfluid flow in the ring should be quite stable²⁴, with $\Omega_c^+ \gg \Omega_0$. Changing U will change the critical angular velocities, Ω_c^\pm , and the size of the hysteresis loop. Rotating the weak link in the azimuthal direction at angular frequency Ω can drive transitions, or phase slips, between the quantized circulation states⁸ (Fig. 2b).

To observe hysteresis in these phase slips, we use a two-step experimental sequence (Fig. 2c). After condensing the atoms into the ring trap, the BEC is prepared in either the $n = 0$ or the $n = 1$ circulation state by either not rotating the weak link or by rotating it at $\Omega_1 = 1.1$ Hz. The fidelity with which this procedure generates the expected initial state is $\approx 97\%$. We then rotate the weak link at various angular velocities, Ω_2 , in the range -0.3 to 1.2 Hz, for an additional 2 s. In step 1, U is ramped to $U_1 \approx 1.1\mu_0$, where μ_0 is the global chemical potential. In step 2, U is ramped to a chosen U_2 . The transitions $n = 0 \rightarrow 1$ and $n = 1 \rightarrow 0$ occur at different values of Ω_2 and form hysteresis loops (Fig. 3a–f). Each plot shows the measured hysteresis loop for a specific U_2 value. As U_2 is increased, both Ω_c^+ and Ω_c^- approach $\Omega_0/2$; that is, the hysteresis loop becomes smaller. The observed transitions are not sharp, unlike those in Fig. 1b. The dominant broadening mechanism is probably shot-to-shot atom number fluctuations, but the non-zero temperature (~ 100 nK) may also contribute (Supplementary Information).

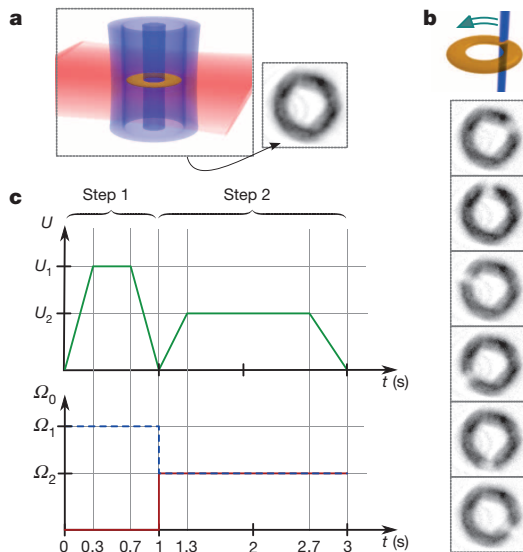


Figure 2 | Experimental set-up and procedure. a, Schematic and *in situ* images of our trap, which is formed by crossing a ring-shaped dipole trap for radial confinement and a sheet trap for vertical confinement. b, Schematic and *in situ* images of a ring rotated by a repulsive weak link. c, Two-step experimental sequence: the height, U , of the repulsive potential and the angular rotation rate, Ω , as a function of time. Step 1 sets the initial winding number using Ω_1 (either 0 or 1.1 Hz) and U_1 ($\sim 1.1\mu_0$); step 2 probes the hysteresis with Ω_2 and U_2 (see text).

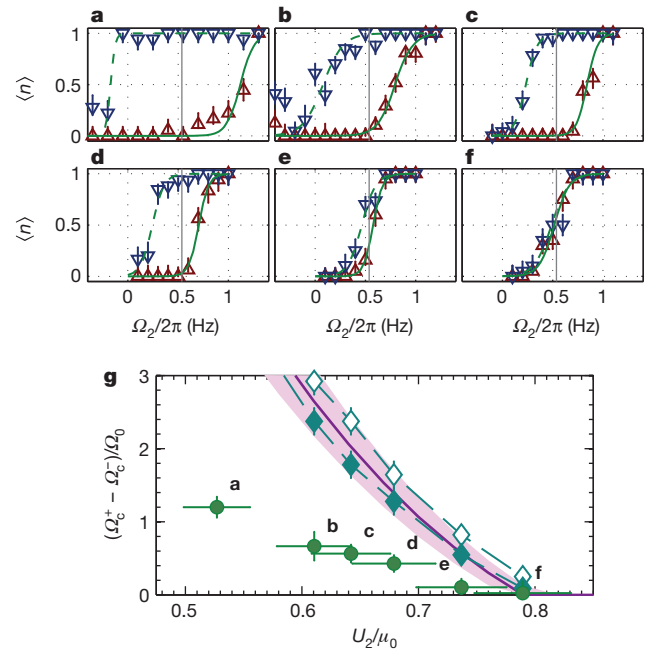


Figure 3 | Hysteresis data. a–f, Hysteresis loops with sigmoid fits. The red up-triangles and blue down-triangles show the winding number n averaged over ~ 20 shots when starting with $n = 0$ and, respectively, $n = 1$. All error bars show the 68% confidence interval. The fits determine Ω_c^\pm and $\Omega_0/2$ (vertical grey lines; Methods) and their uncertainties. g, Hysteresis loop size versus U_2 . The green circles show the experimental data. The magenta line and band are respectively the prediction and uncertainty of an effective one-dimensional hydrodynamic model²⁸. The open and filled cyan diamonds are the results of our GPE simulation with $\Lambda = 0$ and, respectively, $\Lambda = 0.01$.

Figure 3g shows the measured size of the hysteresis loop, $(\Omega_c^+ - \Omega_c^-)/\Omega_0$, as a function of the strength of the weak link; the size of the loop monotonically decreases with increasing U_2/μ_0 until it reaches a value consistent with zero near $U_2/\mu_0 \approx 0.75$. To predict the size of the hysteresis loop, we used two models. First we used an effective one-dimensional model that computes the fluid velocity in the rotating frame as a function of Ω_2 . We assume that Ω_c^\pm will occur when this velocity reaches the local speed of sound²⁸. We also simulated our system with the three-dimensional, time-dependent Gross–Pitaevskii equation (GPE). These two approaches predict hysteresis and are consistent, suggesting that both theories predict that Ω_c^\pm is determined by the sound speed. Despite occurring at the sound speed, the observed excitations in the GPE simulation are vortex–antivortex pairs. Perhaps most strikingly, there is a large discrepancy between our models and experiment (Fig. 3g).

One property of the system that our models fail to include is dissipation. As another approach, we added dissipation to the GPE phenomenologically²⁹, by modifying it as follows:

$$i\hbar \frac{\partial \psi}{\partial t} = (1 - i\Lambda) \left[-\frac{\hbar^2}{2m} \nabla^2 + V(x, y, z, t) + gN|\psi|^2 - \mu \right] \psi$$

Here ψ is the BEC wavefunction, g is the interaction strength, V is the externally applied potential (trap and weak link), N is the atom number, μ is the chemical potential of the initial stationary state and Λ is the dissipation parameter. With $\Lambda = 0.01$, a reasonable value for our experiment, the hysteresis loop size decreases as shown in Fig. 3g but not significantly by comparison with the discrepancy with experiment. Increasing the dissipation parameter does not improve the agreement (Supplementary Information). However, it is clear that dissipation is important. In fact, dissipation is essential and implicitly assumed in the energy landscape picture described in Fig. 1: dissipation allows the system to relax to the minima of the landscape; without dissipation, the system cannot change its energy.

To gain insight into the role of dissipation, we consider a toy model where the relevant excitations are vortex–antivortex pairs, and derive the associated energy landscape. If an antivortex and a vortex were to be nucleated at the inner and, respectively, outer edges (Fig. 4, left inset), move to the centre and annihilate, then the winding number would change by one unit. The energy of a vortex–antivortex pair in a perfectly hard-walled ring trap in the presence of a velocity field has been derived using the method of images³⁰. In the limit that the width of the annulus, d , is much less than R , this energy reduces to

$$E = \pi \rho d R v^2 + 2\pi \frac{\hbar \rho s v}{m} + 2\pi \frac{\rho \hbar^2}{m^2} \ln \left[\frac{d}{\pi \xi} \sin \left(\frac{\pi s}{d} \right) \right] \quad (1)$$

where s is the separation between the vortices, v is the velocity of the superfluid, ρ is the effective two-dimensional mass density, and $\xi = \sqrt{\hbar^2 / 2mgN|\psi|^2}$ is the healing length of the condensate and is therefore the core size. This equation applies to a system with a uniform annulus width, d , and uniform velocity, v . To apply this model to our system, we take d to be the effective width of the annulus in the weak-link region and we take $v = v_m$, the maximum velocity in the weak link. For $v_m = 0$, equation (1) has a maximum at $s = d/2$ and diverges negatively at d and 0 . Such divergence is unphysical, because the non-zero radii of the vortices prevent them from coming arbitrarily close to each other or the wall. We assume that the distance of closest approach to the walls is $C\xi$ and that that between vortices is $2C\xi$, where C is of order unity. Thus, s ranges from $s_{\min} = 2C\xi$ to $s_{\max} = d - 2C\xi$. We assume that vortices annihilate at s_{\min} and enter the annulus at s_{\max} .

We plot the energy landscape described above in Fig. 4 (right inset) for several different v_m values and constant d . The two stable states, at s_{\min} and s_{\max} , represent a winding number difference of one. This implies that for a phase slip to occur the vortex pair must nucleate at either s_{\max} or s_{\min} and move to the opposite extreme. This happens when the energy barrier disappears, that is, when $dE/ds|_{s=s_{\min}} = dE/ds|_{s=s_{\max}} = 0$. This defines the critical velocity, $v_m = v_c$ as

$$v_c = \pm \frac{\pi \hbar}{md} \cot \left(\pi \frac{2C\xi}{d} \right) \quad (2)$$

where the plus and minus signs respectively refer to starting at s_{\min} and s_{\max} .

To compare this model to our experiment, we computed the critical velocity in the weak link from the transitions in the hysteresis loops.

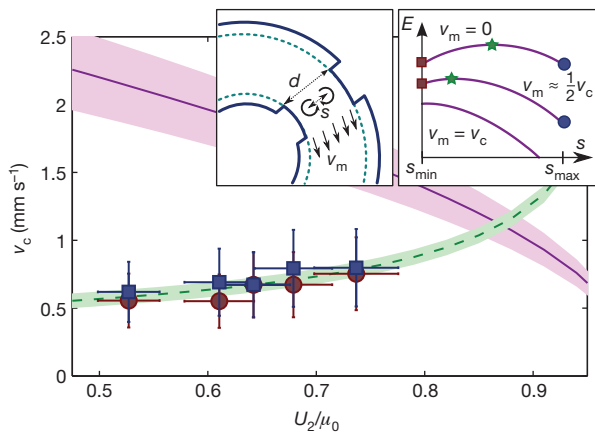


Figure 4 | Extracted critical velocities versus U_2 . The red circles and blue squares show the critical velocities extracted from Ω_c^+ and, respectively, Ω_c^- . The magenta line and band respectively show the estimate and uncertainty of the local speed of sound. The green dashed line and band respectively show the best fit of the toy model of vortex creation and its statistical uncertainty. All uncertainties are 1σ . Left inset, diagram of a vortex–antivortex pair with separation s in a weak link of width d . Right inset, energy landscape as a function of s for three different values of the velocity in the weak-link region, v_m , showing the stable states (squares, circles) and the energy barrier (stars).

The critical velocity is not a simple function of U_2/μ_0 and Ω_2 ; rather, the requirements of quantized winding number and, in a frame co-rotating with the weak link, continuity of flow, require a self-consistent solution for the flow velocity around the entire ring (Methods). Figure 4 shows the result of this calculation. In direct contrast to the local speed of sound in the weak link, which decreases as $\sqrt{1 - U_2/\mu_0}$, we find that the critical velocity increases. The observed critical velocity is well fitted to equation (2) with a single value of $2C\xi/d$. This value implies a distance of closest approach $2C\xi \approx 0.4d$. Over the region of interest, the value of C ranges from 1.5 to 0.7, agreeing with the assumption that it is of order unity (C is calculated using the best estimates of ξ and d , both of which vary with U). The fact that the data can be fitted using this crude model suggests that vortices are the relevant excitations and that equation (1) (or something that captures similar physics) gives a good prediction of the energy landscape.

Our hysteretic system has the essential features of the radio-frequency superconducting quantum interference device (SQUID). Just as SQUIDS detect magnetic fields, our analogous system can detect rotations. Hysteresis has an important role in radio-frequency SQUIDS, where it is used as a readout mechanism. In our system, hysteresis will also be important, allowing for greater accuracy by cancelling systematic effects. The hysteresis loops are centred on $\pm \Omega_0/2$ for different directions of rotation; therefore, we can measure the asymmetry in the measurements of $\pm \Omega_0/2$ to extract an unknown bias rotation. Such measurements may cancel out effects such as asymmetries in the ring potential.

We have measured hysteresis in a dilute atomic-gas BEC, a phenomenon that is as fundamental to superfluidity as the existence of persistent currents and critical velocities. Our studies suggest that the elementary excitations involved in hysteresis are vortices and that dissipation has an important role in the dynamics. We suspect that more sophisticated models that include dissipation will yield better agreement. Finally, beyond being an atomtronic rotation sensor, it is possible that in the hysteretic regime this device could act as classical memory or a digital noise filter in future atomtronic circuits.

METHODS SUMMARY

The ring-shaped BEC, which contains approximately 4×10^5 ^{23}Na atoms, is created from a cloud of laser-cooled atoms by evaporation, first in a magnetic trap and then in a ring-shaped optical dipole trap. The optical dipole trap is shaped roughly like a flattened torus (Methods), with measured harmonic trap frequencies of 472(4) Hz in the vertical direction and 188(3) Hz in the radial direction. (All uncertainties are the uncorrelated combination of 1σ statistical and systematic uncertainties unless stated otherwise.) The mean radius of the trap is 19.5(4) μm . The weak link is created by a blue-detuned laser beam (Methods). Time-of-flight expansion of the condensate allows us to determine the winding number by measuring the size of the central hole that appears in the cloud.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 September; accepted 12 December 2013.

- Pepino, R., Cooper, J., Anderson, D. & Holland, M. Atomtronic circuits of diodes and transistors. *Phys. Rev. Lett.* **103**, 140405 (2009).
- Beeler, M. C. *et al.* The spin Hall effect in a quantum gas. *Nature* **498**, 201–204 (2013).
- Silver, A. H. & Zimmerman, J. E. Quantum states and transitions in weakly connected superconducting rings. *Phys. Rev.* **157**, 317–341 (1967).
- Zimmerman, J. E. Design and operation of stable rf-biased superconducting point-contact quantum devices, and a note on the properties of perfectly clean metal contacts. *J. Appl. Phys.* **41**, 1572–1580 (1970).
- Mueller, E. J. Superfluidity and mean-field energy loops: hysteretic behavior in Bose-Einstein condensates. *Phys. Rev. A* **66**, 063603 (2002).
- Ramanathan, A. *et al.* Superflow in a toroidal Bose-Einstein condensate: an atom circuit with a tunable weak link. *Phys. Rev. Lett.* **106**, 130401 (2011).
- Moulder, S., Beattie, S., Smith, R. P., Tammuz, N. & Hadzibabic, Z. Quantized supercurrent decay in an annular Bose-Einstein condensate. *Phys. Rev. A* **86**, 013629 (2012).
- Wright, K. C., Blakestad, R. B., Lobb, C. J., Phillips, W. D. & Campbell, G. K. Driving phase slips in a superfluid atom circuit with a rotating weak link. *Phys. Rev. Lett.* **110**, 025302 (2013).
- Onofrio, R., Raman, C. & Vogels, J. Observation of superfluid flow in a Bose-Einstein condensed gas. *Phys. Rev. Lett.* **85**, 2228–2231 (2000).

10. Inouye, S. *et al.* Observation of vortex phase singularities in Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 080402 (2001).
11. Engels, P. & Atherton, C. Stationary and nonstationary fluid flow of a Bose-Einstein condensate through a penetrable barrier. *Phys. Rev. Lett.* **99**, 160405 (2007).
12. Miller, D. *et al.* Critical velocity for superfluid flow across the BEC-BCS crossover. *Phys. Rev. Lett.* **99**, 070402 (2007).
13. Neely, T. W., Samson, E. C., Bradley, S., Davis, M. J. & Anderson, B. P. Observation of vortex dipoles in an oblate Bose-Einstein condensate. *Phys. Rev. Lett.* **104**, 160401 (2010).
14. Desbuquois, R. *et al.* Superfluid behaviour of a two-dimensional Bose gas. *Nature Phys.* **8**, 645–648 (2012).
15. Albiez, M. *et al.* Direct observation of tunneling and nonlinear self-trapping in a single bosonic Josephson junction. *Phys. Rev. Lett.* **95**, 010402 (2005).
16. Levy, S., Lahoud, E., Shomroni, I. & Steinhauer, J. The a.c. and d.c. Josephson effects in a Bose-Einstein condensate. *Nature* **449**, 579–583 (2007).
17. Diakonov, D., Jensen, L., Pethick, C. & Smith, H. Loop structure of the lowest Bloch band for a Bose-Einstein condensate. *Phys. Rev. A* **66**, 013604 (2002).
18. Watanabe, G., Yoon, S. & Dalfovo, F. Swallowtail band structure of the superfluid Fermi gas in an optical lattice. *Phys. Rev. Lett.* **107**, 270404 (2011).
19. Morsch, O. & Oberthaler, M. Dynamics of Bose-Einstein condensates in optical lattices. *Rev. Mod. Phys.* **78**, 179–215 (2006).
20. Kojima, H., Veith, W., Putterman, S., Guyon, E. & Rudnick, I. Vortex-free Landau state in rotating superfluid helium. *Phys. Rev. Lett.* **27**, 714–718 (1971).
21. Schwab, K., Bruckner, N. & Packard, R. Detection of the Earth's rotation using superfluid phase coherence. *Nature* **386**, 585–587 (1997).
22. Schwab, K., Bruckner, N. & Packard, R. The superfluid ^4He analog of the RF SQUID. *J. Low Temp. Phys.* **110**, 1043–1104 (1998).
23. Plischke, M. & Bergersen, B. *Equilibrium Statistical Physics* 3rd edn, 83–90 (World Scientific, 2006).
24. Baharian, S. & Baym, G. Bose-Einstein condensates in toroidal traps: instabilities, swallowtail loops, and self-trapping. *Phys. Rev. A* **87**, 013619 (2013).
25. Recati, A., Zambelli, F. & Stringari, S. Overcritical rotation of a trapped Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 377–380 (2001).
26. Madison, K., Chevy, F., Bretin, V. & Dalibard, J. stationary states of a rotating Bose-Einstein condensate: routes to vortex nucleation. *Phys. Rev. Lett.* **86**, 4443–4446 (2001).
27. Hoskinson, E., Sato, Y., Hahn, I. & Packard, R. E. Transition from phase slips to the Josephson effect in a superfluid ^4He weak link. *Nature Phys.* **2**, 23–26 (2006).
28. Watanabe, G., Dalfovo, F., Piazza, F., Pitaevskii, L. P. & Stringari, S. Critical velocity of superfluid flow through single-barrier and periodic potentials. *Phys. Rev. A* **80**, 053602 (2009).
29. Choi, S., Morgan, S. & Burnett, K. Phenomenological damping in trapped atomic Bose-Einstein condensates. *Phys. Rev. A* **57**, 4057–4060 (1998).
30. Fetter, A. L. Low-lying superfluid states in a rotating annulus. *Phys. Rev.* **153**, 285–296 (1967).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was partially supported by ONR, the ARO atomtronics MURI, and the NSF through the PFC at the JQI and grant PHY-1068761. S.E. is supported by a National Research Council postdoctoral fellowship. We wish to thank K. Wright, W. T. Hill III and A. Kumar for valuable discussions and experimental assistance.

Author Contributions S.E., J.G.L. and F.J. took the experimental data. N.M., C.W.C. and M.E. developed and performed the GPE simulations. All authors were involved in analysis and discussions of the results, and contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.K.C. (gretchen.campbell@nist.gov).

METHODS

Optical dipole traps. Our optical dipole trap is formed by the combination of two laser beams. A blue-detuned ($\lambda = 532$ nm) laser beam passes through a ring-shaped intensity mask, and the shadow is imaged onto the atoms, forming a repulsive, ring-shaped potential. This trap combines with an attractive confining potential in the vertical direction, generated by a red-detuned ($\lambda = 1,064$ nm) laser beam shaped like a sheet. If the imaging resolution were perfect, the trap would be hard-walled in the radial direction, but it is in fact closer to a Gaussian with a $1/e^2$ radius of $8.9(9)$ μm .

To create the weak link, we use an acoustic optical deflector to elongate a blue-detuned, focused, Gaussian beam by scanning radially at 2 kHz. The beam is turned on and off with a 300-ms linear ramp. The $1/e^2$ half-width of the weak link along the azimuthal direction is approximately 6 μm , limited by the resolution of our imaging system. The size of the weak link along the radial direction is $\sim 50\%$ larger than the Thomas–Fermi width of the BEC.

We calibrated the weak link by observing the atomic density depletion caused by the weak-link potential (Supplementary Information). The dominant uncertainty in U_2/μ_0 is in the common calibration and is reflected in the horizontal error bars in Figs 3 and 4; the relative uncertainties between the points are smaller.

BEC parameters. Approximately 4×10^5 atoms comprise the BEC after evaporation first in a magnetic time-orbiting potential trap and subsequently in the optical traps described above. We estimate the global chemical potential, μ_0 , to be $\mu_0/\hbar \approx 2\pi \times 1.7$ kHz and the corresponding Thomas–Fermi full-widths in the vertical and radial directions to be 3.2 and 8.1 μm , respectively. Given this mean radius, we expect that $\Omega_0 = 1.19(4)$ Hz, in rough agreement with the measured value of $\Omega_0 = 1.05(5)$ Hz (by assuming the hysteresis loops are centred on $\Omega_0/2$).

Measurement of the winding number. To measure the final rotational state after rotating, the BEC is released from the trap and imaged after a 10-ms time of flight. As the BEC expands, rotation will cause a hole to appear in the centre. As with the winding number, the size of this feature is quantized and enables determination of the final circulation state^{7,8}. Direct release from the repulsive dipole trap does not allow the hole to be resolved during the 10-ms time of flight, so we transfer first to an attractive ring and apply a decompression procedure similar to that in ref. 8, before using partial transfer absorption imaging³¹.

Estimating the uncertainty in the average winding number. Given that the outcome of any given experiment is either $n = 0$ or $n = 1$, traditional methods of estimating the uncertainty in the mean value (for example Gaussian statistics) are not applicable. The uncertainty in this average can be estimated by the cumulative beta distribution, which is appropriate for experiments that yield binary results³².

Fitting the hysteretic transitions and determining Ω_c^\pm and its uncertainty. We use a sigmoid of the form $1/(\exp[-(\Omega_2 - \Omega_1)/\delta\Omega] + 1)$ to fit the data as in Fig. 3, where Ω_1 and $\delta\Omega$ are the fit parameters. Although this fit well describes the data, the relationship between Ω_c^\pm and the fit parameters depends on the mechanisms for the broadening of the transition. For example, consider a model where thermal fluctuations drive the system over the energy barrier. This would occur when the energy barrier becomes of the order of $k_B T$, where k_B is the Boltzmann constant and T is the temperature. The dynamics of this process is random and would lead to phase slips at lower values of Ω_2 for the $0 \rightarrow 1$ transition (and higher values for

$1 \rightarrow 0$ transitions). In principle, this effect would cause a broadening of the transition region, and the zero-temperature Ω_c^\pm would then correspond to the value of Ω_2 where probability for a transition equals unity. However, a different mechanism could be responsible for the broadening. In particular, atom number fluctuations can change U_2/μ_0 and, therefore, Ω_c^\pm , from shot to shot. On average, this leads to a broadening. On the basis of the experimentally observed change in Ω_c^\pm versus the strength of the weak link, and our atom number shot-to-shot fluctuations of $\sim 16\%$ (which represent the peak-to-peak fluctuations for 95% of the data), we expect the transitions to be approximately 0.12 Hz wide, compared with the average of 0.18 Hz. Because atom number fluctuations explain most of the width, we take $\Omega_c^\pm = \Omega_1$, and, to account for the possibility of finite-temperature or other unknown broadening effects, take the 1σ uncertainty to be $(3/2)\delta\Omega$.

Extracting the critical velocity. Extracting the critical velocity in our system is non-trivial because the flow must satisfy the requirements of quantized winding number and continuity of flow in the frame rotating with the weak link. (Continuity of flow does not occur in any other frame.) One counterintuitive result of these requirements is that moving the weak link will impart some angular momentum to the superfluid as viewed from the fixed, laboratory frame, even in the $n = 0$ state^{20,33}. To extract the critical velocity given these constraints, we work in the rotating frame. The velocity, v_r , of atoms in the rotating frame is related to the rotation rate of the weak link by

$$\frac{m}{\hbar} \int_0^{2\pi} v_r(\theta) R d\theta + 2\pi \frac{\Omega}{\Omega_0} = 2\pi n \quad (3)$$

where θ is the azimuthal angle. This equation is an expression of the Bohr–Sommerfeld quantization condition. The first term represents the phase accumulated by the atoms after integrating once around the ring, and the second term represents the Sagnac phase that appears as a result of transforming into the rotating frame. In the rotating frame, the atom velocity, $v_r(\theta)$, and the mass density, $\rho(\theta)$, satisfy a continuity equation, $\rho(\theta_1)v_r(\theta_1) = \rho(\theta_2)v_r(\theta_2)$, where θ_1 and θ_2 are any two azimuthal angles. Given U_2/μ_0 , we determine the equivalent one-dimensional density, $\rho(\theta)$, by integrating over the radial and vertical directions of our cloud using the Thomas–Fermi approximation. For a given rotation rate, Ω , and density, $\rho(\theta)$, equation (3) and the continuity equation uniquely determine $v_r(\theta)$ and, in particular, the velocity in the weak link, $v_m = \max[v_r(\theta)]$. The critical velocity is then taken to be the value of v_m when the weak link is rotated at the critical rotation rate Ω_c^\pm .

When $\Omega_c^+ - \Omega_c^- \rightarrow 0$, this method of extracting the critical velocity is unreliable and thus we neglect the point near $U \approx 0.8\mu_0$ (Supplementary Information). Going slightly farther into the regime where $U > \mu_0$ results in the BEC becoming simply connected.

31. Ramanathan, A. *et al.* Partial-transfer absorption imaging: a versatile technique for optimal imaging of ultracold gases. *Rev. Sci. Instrum.* **83**, 083119 (2012).
32. Cameron, E. On the estimation of confidence intervals for binomial population proportions in astronomy: the simplicity and superiority of the Bayesian approach. *Publ. Astron. Soc. Aust.* **28**, 128–139 (2011).
33. Lamb, H. *Hydrodynamics* 6th edn, 123–125 (Dover, 1932).

Quantum error correction in a solid-state hybrid spin register

G. Waldherr^{1*}, Y. Wang^{1*}, S. Zaiser¹, M. Jamali¹, T. Schulte-Herbrüggen², H. Abe³, T. Ohshima³, J. Isoya⁴, J. F. Du⁵, P. Neumann¹ & J. Wrachtrup^{1,6}

Error correction is important in classical and quantum computation. Decoherence caused by the inevitable interaction of quantum bits with their environment leads to dephasing or even relaxation. Correction of the concomitant errors is therefore a fundamental requirement for scalable quantum computation^{1–7}. Although algorithms for error correction have been known for some time, experimental realizations are scarce^{2–7}. Here we show quantum error correction in a heterogeneous, solid-state spin system^{8–21}. We demonstrate that joint initialization, projective readout and fast local and non-local gate operations can all be achieved in diamond spin systems, even under ambient conditions. High-fidelity initialization of a whole spin register (99 per cent) and single-shot readout of multiple individual nuclear spins are achieved by using the ancillary electron spin of a nitrogen–vacancy defect. Implementation of a novel non-local gate generic to our electron–nuclear quantum register allows the preparation of entangled states of three nuclear spins, with fidelities exceeding 85 per cent. With these techniques, we demonstrate three-qubit phase-flip error correction. Using optimal control, all of the above operations achieve fidelities approaching those needed for fault-tolerant quantum operation, thus paving the way to large-scale quantum computation. Besides their use with diamond spin systems, our techniques can be used to improve scaling of quantum networks relying on phosphorus in silicon¹⁹, quantum dots²², silicon carbide¹¹ or rare-earth ions in solids^{20,21}.

Heterogeneous spin systems for quantum computation seek to combine the strengths of their constituents: electron spins for fast, versatile and high-fidelity readout and control^{8–17,19,20}, and nuclear spins operating as well-shielded quantum bits yielding record values for coherence times (in, for example, diamond and silicon)^{9,18–20}. However, even for such nuclear spins, unavoidable interaction with their environment (including the readout electron spin) occurs, leading to decoherence, which inevitably results in errors. Even small changes of a quantum state lead to finite error probabilities, unlike classical computing, where small state changes (for example, voltage) do not affect the final result. The threshold theorem states that the final error probability of a quantum algorithm on a large register can be made arbitrarily small if the probability of an error on a single qubit is below a certain threshold, and quantum error correction (QEC) is applied¹. Despite being a cornerstone of scalable quantum systems, QEC has only been implemented on one solid-state system so far, namely superconducting circuits⁷. This is because of the challenging demands of high-fidelity non-local control of at least three qubits.

Here we demonstrate a high degree of control sufficient for QEC on a hybrid spin system, namely the nitrogen–vacancy defect (NV) in diamond^{10,13–15,17,18,23–25}. Nuclear spins randomly distributed within an isotopically purified diamond with $\sim 0.2\%$ ^{13}C (see Methods) are addressed, read out and initialized by the electron spin of the defect. The electron spin in turn is polarized and read out by optical means.

Specifically, we use three nuclear spins (^{14}N with $I = 1$ and two ^{13}C , that is, $^{13}\text{C}_1$ and $^{13}\text{C}_2$, with $I = 1/2$) as quantum bits (Fig. 1a). All nuclear spins of the register can be individually addressed via their hyperfine interaction with the electron spin, which shows spectral signatures of the 12 register states (Fig. 1b). We note that during revision of this manuscript, we became aware of a preprint²⁶ in which QEC was implemented with two nuclear spins and one electron spin.

Projective, single-shot readout of the nuclear spins is essential for high-fidelity initialization, proving genuine quantum correlations, and fast measurements (here a factor of ~ 5 faster than conventional readout). The method correlates the electron spin with the nuclear qubit state by a nuclear state specific controlled electron NOT (C_nNOT_e) gate^{10,27}. The corresponding measurement sequence is shown in Fig. 1c. Here we have extended this method to three nuclei. The non-destructive nature of the readout process is ensured by application of a magnetic field^{10,27}. As is apparent from Fig. 1e, the lifetime of the nuclear spin states is, indeed, much longer than the time needed to measure their states. To further quantify the fidelity of our readout, Fig. 1e also shows histograms of measurement results for nuclear spin readout. By defining a threshold between the two maxima of the distributions, the state of the spin is determined by measuring whether the photon count is below or above this threshold. Here we achieve readout fidelities of 95.8% for ^{14}N , 96.9% for $^{13}\text{C}_1$ and 99.6% for $^{13}\text{C}_2$.

To initialize the nuclear register, we first apply polarization transfer from the electron spin²³, and then perform projective readout of the register (Fig. 2a). By this approach, we achieve high-fidelity initialization, while at the same time overcoming the limitation of purely measurement-based initialization where the success rate scales inversely with the number of register states. With a combination of these two methods, the initialization fidelity reaches values of up to 99% for the whole register into state $|^{14}\text{N}, ^{13}\text{C}_1, ^{13}\text{C}_2\rangle = |0, 0, 0\rangle$ (see Methods). Note that the electron spin is not explicitly included as a quantum bit in the register, but is solely used for the control of nuclear qubits, thus fully exploiting the advantages of our hybrid quantum system. Additionally, initialization of the NV into its negative charge state is verified by charge state post-selection (see Methods).

Previously reported entanglement of two nuclear spins relied on pseudo-initialization²⁴, or non-deterministic entanglement-by-measurement¹⁴ at cryogenic temperatures. Instead, we use a novel scheme to create nuclear entanglement, which is based on a conditional π -phase (CPhase) gate on the electron spin³¹. Such a phase gate can either be implemented by adiabatic, geometric phase gates, which offer intrinsic robustness, or by exploiting the π phase shift that is acquired by a full 2π state rotation. Here we use the latter, in combination with optimal control, to achieve high-fidelity implementation, which is both robust and fast. Optimal control offers a general framework for creating robust state manipulation by a series of control steps (here with varying amplitude and phase of the control field, similar to dynamical decoupling¹³),

¹Physikalisches Institut und Research Center SCOPE, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany. ²Department of Chemistry, Technical University of Munich, 85747 Garching, Germany. ³Japan Atomic Energy Agency, Takasaki, Gunma 370-1292, Japan. ⁴Research Center for Knowledge Communities, University of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan. ⁵Hefei National Laboratory for Physical Sciences at the Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China. ⁶Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany.

*These authors contributed equally to this work.

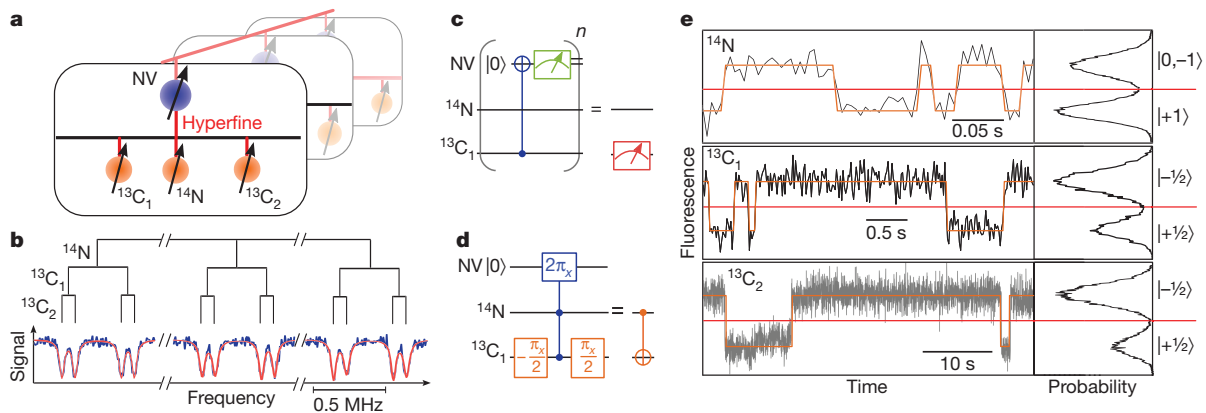


Figure 1 | Single-shot readout and control of a nuclear spin register via the NV. **a**, Illustration of the quantum register consisting of three nuclear spins (orange circles with black arrows) coupled by hyperfine interaction to the NV electron spin (blue circles with black arrows), forming a prototype single node of a larger register^{15,17}. $^{13}\text{C}_{1,2}$ refers to ^{13}C nuclei at different lattice positions. **b**, Hyperfine splitting (^{14}N , 2.16 MHz; $^{13}\text{C}_1$, 413 kHz; $^{13}\text{C}_2$, 89 kHz) of the $m_s = 0$ to -1 transition into 12 individual lines. **c**, 'Wire diagram' for projective

readout of nuclear spins^{10,27}; the diagram here is for a ^{13}C spin. **d**, Wire diagram of a CNOT gate between two nuclear spins. **e**, Left side, time trace of NV fluorescence photons during repetitive readout of a particular nuclear spin (see **c**). The timescales of the dynamics depend on the nuclear spin life times and the C_nNOT_e gate durations. Right side, photon-counting histograms of the time traces. The red line indicates the threshold for nuclear spin state discrimination.

which are numerically optimized and therefore suitable for complex Hamiltonians²⁸ (see Methods). Our approach is robust against the fast decoherence of the electron spin, while still exploiting hyperfine interaction, which allows for ultra-fast control of nuclear spins. It thus makes best use of the interactions in our electron nuclear hybrid spin quantum systems. Specifically, a robust, controlled $2\pi_{x,y}$ rotation (see Methods for nomenclature) of the electron spin conditional on the state of the nuclear register induces a π phase shift of the corresponding register state (for example, $|11\rangle \rightarrow -|11\rangle$), resulting in a nuclear CNOT gate (Fig. 1d).

For three qubits, there are two types of fully entangled state that cannot be transformed into each other by local gates, the GHZ (Greenberger–Horne–Zeilinger) state ($|\text{GHZ}\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$) and the W state ($|\text{W}\rangle = (|011\rangle + |101\rangle + |110\rangle)/\sqrt{3}$)²⁹. Whereas the GHZ state decoheres on single qubit dephasing, this is not the case for the W state, as loss of coherence of a single qubit leaves a fully coherent state of the remaining spins. Using our CPhase gate, we create a GHZ-like entangled state (Fig. 2c, e) and a W state (Fig. 2d, f) of the three nuclear spins. The resulting density matrices obtained by state tomography (see Methods) are shown in Fig. 2e, f. The estimated fidelity $F = \langle \psi | \rho | \psi \rangle$ of the final state ρ is $F = (88 \pm 1)\%$ for the GHZ state and $F = (85 \pm 3)\%$ for the W state, corrected by the readout fidelity of 89%. Factors reducing the

fidelity (apart from the actual entanglement sequence) are T_1 decay of the electron spin ($\sim 96\%$) and initialization and post-selection fidelity ($\sim 97\%$).

To show genuine three-qubit entanglement, we also measure violation of the Mermin inequality³⁰. For the GHZ-like state in Fig. 2c, the inequality is $|\langle \sigma_x \sigma_x \sigma_z \rangle + \langle \sigma_x \sigma_z \sigma_x \rangle + \langle \sigma_y \sigma_x \sigma_x \rangle - \langle \sigma_y \sigma_z \sigma_z \rangle| \leq 2$ (see Methods; σ_i is the i Pauli operator), which holds for any local, realistic theory. This inequality is violated by quantum mechanics, yielding a maximum value of 4. For the present case, we measured a value of 3.258 ± 0.014 , clearly showing the quantum nature of the state.

To demonstrate the functionality of our hybrid quantum register, we perform QEC with the three nuclear spins. Generally, errors can occur in the form of bit flips, which act as $|0\rangle \leftrightarrow |1\rangle$, or phase flips, which act as $|x_+\rangle \leftrightarrow |x_-\rangle$, $|y_+\rangle \leftrightarrow |y_-\rangle$, where $|x_{\pm}\rangle = |0\rangle \pm |1\rangle$, $|y_{\pm}\rangle = |0\rangle \pm i|1\rangle$. It can be seen that phase flips are identical to bit flips, but in a different basis of the qubit. The error correction protocols for bit flips and phase flips thus are essentially identical, only differing by local $\pi_{x,y}/2$ rotations for basis changes. QEC is based on encoding the information of one physical qubit into two states of a multi-qubit register, forming one logical qubit. The number of errors which can be corrected is given by the bit flip distance d of the two logical states. As an example, for three qubits and logical states $|000\rangle$ and $|111\rangle$, three bit flips ($d = 3$) are

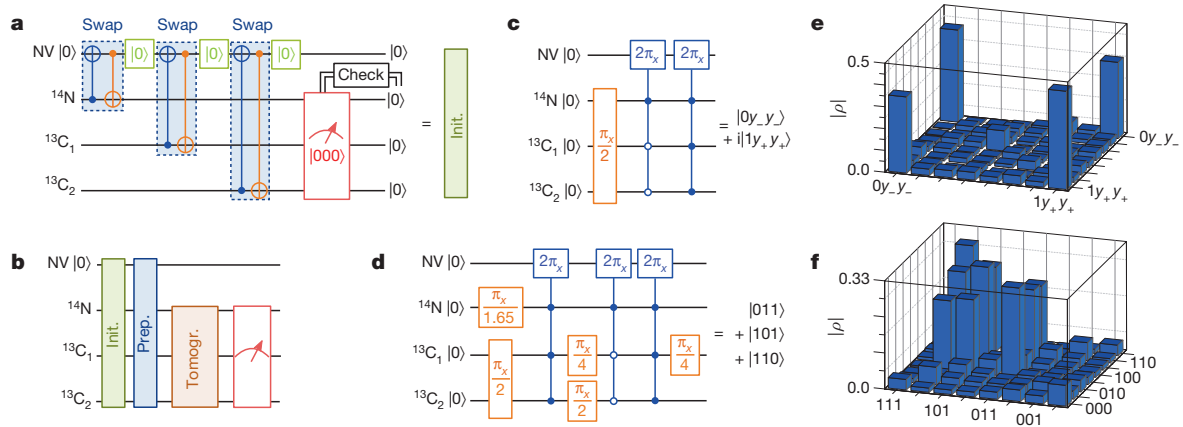


Figure 2 | Three-qubit entangled states. **a**, Initialization (Init.) sequence for the nuclear spin register. The $|0\rangle$ gate indicates optical initialization of the electron spin. Swap-like gates are used to transfer electron spin polarization onto the nuclear spins (Methods). A final measurement probes whether the state of the nuclear spins is $|000\rangle$. **b**, Measurement sequence. The tomography

(Tomogr.) step consists of local gates on the nuclear spins. The preparation (Prep.) step is either the sequence shown in **c** for the GHZ state or the sequence shown in **d** for the W state. **e**, **f**, Reconstructed density matrices (absolute values of ρ) for the GHZ state (**e**) and for the W state (**f**; see text).

necessary to convert one state into the other. Thus, erroneous states resulting from single bit flips ($d = 1$) stay always closer to the original state and can therefore be corrected by a majority vote principle (see below), whereas multiple flips ($d > 1$) will inevitably destroy the information and render the protocol ineffective. Although errors in quantum mechanics appear as continuous state rotations, the subsequent error detection projects these rotations onto the two possible cases of not having an error or having an error on the qubit, with probability given by the angle of rotation. Thus, after detection, possible errors can be reverted to restore the original state.

Here we demonstrate a phase flip correction protocol, as the phase flip error rate is usually much higher than the bit flip error rate. Figure 3a shows the QEC sequence^{6,7}. The code does not need error detection measurements: it corrects and restores the state of one qubit, and transfers any errors of this qubit onto the two other qubits by coherent evolution, as shown below. To re-initialize the register, the concomitant entropy increase on the other qubits is removed, for example by measurement⁶. The original state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ of $^{13}\text{C}_2$, which we use as an example to demonstrate QEC, is encoded as $\alpha|y_+y_+y_+\rangle + \beta|y_-y_-y_-\rangle$ on the full register. After decoding, the state will be $\alpha|000\rangle + \beta|111\rangle$ in the case of no error, and $\alpha|001\rangle + \beta|110\rangle$ in the case of an error on $^{13}\text{C}_2$. The error is detected via the other two qubits, by flipping them conditional on the state of $^{13}\text{C}_2$, yielding $|00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle)$ and $|11\rangle \otimes (\alpha|1\rangle + \beta|0\rangle)$, respectively. Finally, the state of $^{13}\text{C}_2$ is flipped conditional on the state of the two other spins, resulting in $|00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle)$ if there was no

error, or $|11\rangle \otimes (\alpha|0\rangle + \beta|1\rangle)$ if there was an error, which is, in both cases, the original state $|\psi\rangle$ for $^{13}\text{C}_2$.

To obtain the performance of our implementation, we measure the fidelity of the quantum process on the main qubit (see Methods) as a function of the error probability. Controlled errors on a qubit are introduced by changing the phase of all operations on that qubit after the occurrence of the error. Figure 3b shows the results of our measurements. Indeed, single qubit errors are corrected by the algorithm (green triangles). The added complexity of the correction reduces the fidelity for very low error probability. However, even for simultaneous errors on all qubits (red dots), the fidelity with error correction surpasses the fidelity without error correction (blue circles), if the error probability increases. This region is illustrated by the red area in Fig. 3b.

Our experiments demonstrate a fully functional hybrid spin register. We have combined optimal-control-based error avoidance with error correction with high-fidelity state initialization. A key ingredient of our approach is the separation of readout and control (electron spin) and storage quantum bits (nuclear spins). Although we demonstrated fast and high-fidelity entanglement gates on diamond nuclear spins, our results present a benchmark for similar single-spin systems^{11,19,20,22}. The latter only differ in electron spin readout and preparation techniques, whereas both the control methods based on spin resonance and the access to nuclear spins via hyperfine interaction are identical^{19,20}. The present approach makes use of strongly coupled nuclear spins (hyperfine interaction $>$ inhomogeneous broadening), but restricts the number of eligible defect centres. Including weakly coupled nuclear spins (hyperfine interaction $>$ homogeneous broadening) will increase the number of accessible qubits per electron spin and at the same time render every defect centre a potential quantum register, independent of the particular setting of nuclei (see Methods). In addition, using nuclear spins with very weak hyperfine interaction allows for repetitive error correction as quantum information on those spins is preserved while the remaining register is reset¹⁸.

METHODS SUMMARY

Experimental set-up. The measurements were performed with a home-built confocal microscope. A 532-nm diode-pumped solid-state laser is focused via an oil objective onto the NV to excite its electron. A solid immersion lens (SIL) is created in the diamond to improve fluorescence detection. Fluorescence is filtered by a 650-nm long-pass filter and a 50- μm pinhole, and then detected by a single-photon-counting avalanche photodiode. The sample is a type IIa chemical vapour deposition diamond. NVs are formed by electron irradiation and subsequent annealing. The magnetic field is supplied by a permanent magnet attached to a three-dimensional (3D) positioning system, to align the magnetic field parallel to the NV axis. Here, fields of $\sim 6,200$ G were applied. Electron and nuclear spins are manipulated by radio frequency and microwave signals generated by an arbitrary-waveform generator, which allows for rotations of the spins around an axis in the x - y plane of the Bloch sphere. The actual rotation axis is defined by the signal phase. The total operation in this work is denoted by ϕ_a , where ϕ is the rotation angle and a the rotation axis. **Non-local gates.** Non-local gates are applied via the electron spin, which can be initialized by laser illumination. As can be seen in Fig. 1b, the microwave transition frequency of the electron spin is distinct for the 12 possible states of the nuclear register. By applying frequency-selective microwave pulses, operations on the electron spin can be carried out conditional on any state (or any combination of states) of the register. A $C_n\text{NOT}_e$ is implemented by a π rotation of the electron spin, conditioned on the desired nuclear spin states. Here, the π rotation commences around the x axis of the Bloch sphere if not otherwise indicated.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 September; accepted 27 November 2013.

Published online 29 January; corrected online 12 February 2014 (see full-text HTML version for details).

- Shor, P. W. in *Proc. 37th Symp. Foundations Comput.* 56–65 (IEEE Comp. Soc. Press, 1996); available at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=548464.
- Cory, D. G. et al. Experimental quantum error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).

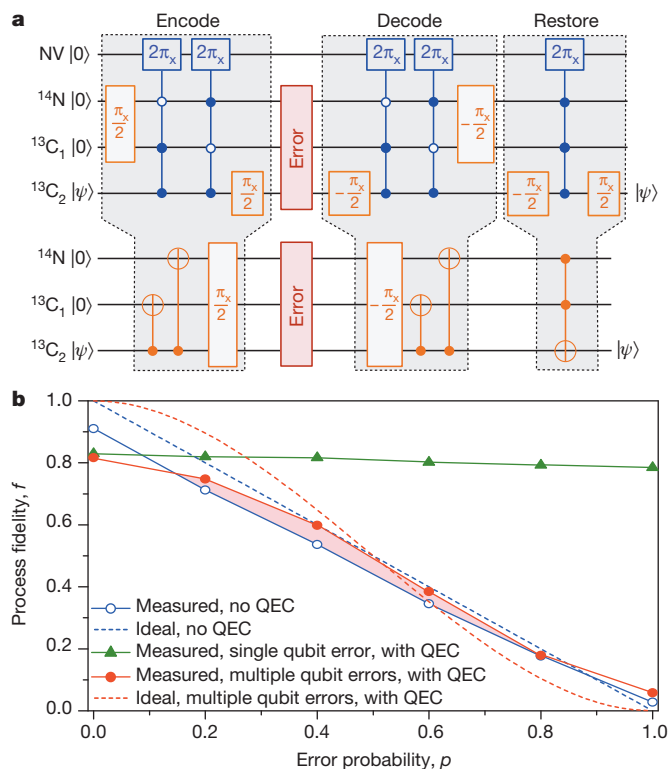


Figure 3 | Quantum error correction. **a**, Quantum circuit for phase flip QEC. Upper part, gates as implemented in the experiment; lower part, effective gates on the nuclear spins. For conditional gates, dots and circles correspond to $|1\rangle$ and $|0\rangle$ states, respectively. **b**, Dependence of measured process fidelity f on the error probability p . The blue circles and the dashed blue line show data obtained without QEC, for our measurements and for the theoretical ideal case ($f(p) = 1 - p$), respectively. For all measurements displayed here, the total duration of the sequence was the same, for better comparison. The QEC code is implemented for errors induced only on $^{13}\text{C}_2$ (green triangles), and for errors with equal probability on all qubits (red dots; dashed red line for the theoretical ideal case with $f(p) = 1 - 3p^2 + 2p^3$). See main text for explanation of red shaded area.

3. Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five-qubit error correcting code. *Phys. Rev. Lett.* **86**, 5811–5814 (2001).
4. Boulant, N., Viola, L., Fortunato, E. & Cory, D. Experimental implementation of a concatenated quantum error-correcting code. *Phys. Rev. Lett.* **94**, 130501 (2005).
5. Moussa, O., Baugh, J., Ryan, C. A. & Laflamme, R. Demonstration of sufficient control for two rounds of quantum error correction in a solid state ensemble quantum information processor. *Phys. Rev. Lett.* **107**, 160501 (2011).
6. Schindler, P. *et al.* Experimental repetitive quantum error correction. *Science* **332**, 1059–1061 (2011).
7. Reed, M. D. *et al.* Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).
8. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
9. Morton, J. J. L. *et al.* Solid-state quantum memory using the ^{31}P nuclear spin. *Nature* **455**, 1085–1088 (2008).
10. Neumann, P. *et al.* Single-shot readout of a single nuclear spin. *Science* **329**, 542–544 (2010).
11. Koehl, W. F., Buckley, B. B., Heremans, F. J., Calusine, G. & Awschalom, D. D. Room temperature coherent control of defect spin qubits in silicon carbide. *Nature* **479**, 84–87 (2011).
12. Pla, J. J. *et al.* A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).
13. Taminiau, T. *et al.* Detection and control of individual nuclear spins using a weakly coupled electron spin. *Phys. Rev. Lett.* **109**, 137602 (2012).
14. Pfaff, W. *et al.* Demonstration of entanglement-by-measurement of solid-state qubits. *Nature Phys.* **9**, 29–33 (2012).
15. Dolde, F. *et al.* Room-temperature entanglement between single defect spins in diamond. *Nature Phys.* **9**, 139–143 (2013).
16. Chekhovich, E. A. *et al.* Nuclear spin effects in semiconductor quantum dots. *Nature Mater.* **12**, 494–504 (2013).
17. Bernien, H. *et al.* Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497**, 86–90 (2013).
18. Maurer, P. C. *et al.* Room-temperature quantum bit memory exceeding one second. *Science* **336**, 1283–1286 (2012).
19. Pla, J. J. *et al.* High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature* **496**, 334–338 (2013).
20. Yin, C. *et al.* Optical addressing of an individual erbium ion in silicon. *Nature* **497**, 91–94 (2013).
21. Kolesov, R. *et al.* Optical detection of a single rare-earth ion in a crystal. *Nature Commun.* **3**, 1029 (2012).
22. Le Gall, C., Brunetti, A., Boukari, H. & Besombes, L. Optical Stark effect and dressed exciton states in a Mn-doped CdTe quantum dot. *Phys. Rev. Lett.* **107**, 057401 (2011).
23. Dutt, M. *et al.* Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science* **316**, 1312–1316 (2007).
24. Neumann, P. *et al.* Multipartite entanglement among single spins in diamond. *Science* **320**, 1326–1329 (2008).
25. Robledo, L. *et al.* High-fidelity projective read-out of a solid-state spin quantum register. *Nature* **477**, 574–578 (2011).
26. Taminiau, T. H., Cramer, J., van der Sar, T., Dobrovitski, V. V. & Hanson, R. Universal control and error correction in multi-qubit spin registers in diamond. Preprint at <http://arxiv.org/abs/1309.5452> (2013).
27. Dréau, A., Spinicelli, P., Maze, J. R., Roch, J.-F. & Jacques, V. Single-shot readout of multiple nuclear spin qubits in diamond under ambient conditions. *Phys. Rev. Lett.* **110**, 060502 (2013).
28. Machnes, S. *et al.* Comparing, optimizing, and benchmarking quantum-control algorithms in a unifying programming framework. *Phys. Rev. A* **84**, 022305 (2011).
29. Dür, W., Vidal, G. & Cirac, J. I. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A* **62**, 062314 (2000).
30. Mermin, N. D. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.* **65**, 1838–1840 (1990).
31. Filidou, V. *et al.* Ultrafast entangling gates between nuclear spins using photo-excited triplet states. *Nature Phys.* **8**, 596–600 (2012).

Acknowledgements We thank F. Dolde, I. Jakobi, M. Kleinmann, F. Jelezko, J. Honert, A. Brunner and C. Walter for experimental help and discussions. We acknowledge financial support from the Max Planck Society, the ERC project SQUITEC, the DFG SFB/TR21, the EU projects DIAMANT, SIQS, QESSENCE and QINVC, the JST-DFG (FOR1482 and FOR1493), and the Volkswagenstiftung.

Author Contributions Y.W., G.W., P.N. and J.W. conceived the experiments; G.W. and S.Z. prepared the sample and performed the experiments; Y.W. calculated the robust pulses; Y.W., G.W. and S.Z. analysed the data; J.J., H.A., T.O. and P.N. performed the electron irradiation; M.J. fabricated the SIL; G.W., J.W., P.N., Y.W., T.S.-H. and J.F.D. wrote the manuscript; and P.N. and J.W. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W. (g.waldherr@physik.uni-stuttgart.de).

METHODS

Experimental set-up. The measurements were performed with a home-built confocal microscope. A 532-nm diode-pumped solid-state laser is focused via an oil objective onto the NV to excite its electron. A solid immersion lens (SIL) was created in the diamond to improve fluorescence detection. Fluorescence is filtered by a 650-nm long-pass filter and a 50 μm pinhole, and then detected by a single photon counting avalanche photodiode. The sample is a type IIa chemical vapour deposition diamond. NVs are formed by electron irradiation and subsequent annealing. The magnetic field is supplied by a permanent magnet attached to a 3D positioning system, to align the magnetic field parallel to the NV axis. Here, fields of $\sim 6,200$ G were applied. Electron and nuclear spins are manipulated by radio-frequency and microwave signals generated by an arbitrary-waveform generator, which allows for rotations of the spins around an axis in the x - y plane of the Bloch sphere. The actual rotation axis is defined by the signal phase. The total operation in this work is denoted by ϕ_a , where ϕ is the rotation angle and a the rotation axis. Additionally, microwave signals are mixed with a high-frequency source to obtain high-frequency arbitrary waveforms at around 15 GHz. Signals are applied via a coplanar waveguide microstructure, which is built directly onto the diamond near the NV, such that optical access to the NV is possible through the gap of the waveguide.

Non-local gates. Non-local gates are applied via the electron spin, which can be initialized by laser illumination. As can be seen in Fig. 1b, the microwave transition frequency of the electron spin is distinct for the 12 possible states of the nuclear register. By applying frequency-selective microwave pulses, operations on the electron spin can be carried out conditional on any state (or any combination of states) of the register. A $C_n\text{NOT}_e$ is implemented by a π rotation of the electron spin, conditioned on the desired nuclear spin states. Here, the π rotation commences around the x axis of the Bloch sphere if not otherwise indicated.

In the following, we illustrate the mechanism for the creation of a two-qubit GHZ-like state using a nuclear CPhase gate based on a conditional $2\pi_x$ rotation of the electron spin. We start with both nuclear qubits initialized in state $|00\rangle$, and apply $\pi_x/2$ rotations on both, to end up in (excluding normalization):

$$|00\rangle - i|01\rangle - |10\rangle - i|11\rangle \quad (1)$$

Now, a CPhase conditional on the nuclear spin state being $|11\rangle$ will change the phase of this state to $-|11\rangle$. This yields the entangled GHZ-like state

$$|00\rangle - i|10\rangle - |10\rangle - i|11\rangle = |0y_-\rangle + |1y_+\rangle \quad (2)$$

where $|y_{\pm}\rangle = |0\rangle \pm i|1\rangle$. This method allows the implementation of nuclear CNOT gates (Fig. 1d).

^{13}C hyperfine interactions. To find an NV with two ^{13}C nuclear spins suitable for single-shot readout, we acquired statistics of possible ^{13}C hyperfine splittings for almost 3,300 NVs. For each analysed NV, the EPR spectrum is measured, as shown in Fig. 1b. A multi-peak function fit was used to determine hyperfine interactions larger than approximately 100 kHz, as the line width of the transition is around 50 kHz. These measurements were performed at magnetic fields of around 10 G. However, the magnetic field was not aligned for all NVs, which can slightly alter the measured hyperfine interaction compared to the value at zero field or at small, aligned magnetic fields. The deviation is expected to be below 2.5% (in the worst case of pure dipolar hyperfine interaction, unfavourable ^{13}C position and perpendicular magnetic field). Finally, we can create a hyperfine probability spectrum in the following way: for each observed hyperfine interaction, we take a Gaussian distribution at the centre of the observed value, and a standard deviation according to the fit accuracy. Summing up over all measured values results in the spectrum shown in Extended Data Fig. 1. By trying to perform single-shot readout on these ^{13}C spins, we found that hyperfine splittings (in kHz) of 124, 211, 384, 422, 517 are usable for single-shot readout at magnetic fields of around 0.5 T, with spin lifetimes (in readout steps) comparable to or better than ^{14}N (ref. 10).

Scaling of strongly and weakly coupled ^{13}C spins. The average number of ^{13}C spins available for single-shot readout depends on the external magnetic field²⁷, the concentration of ^{13}C nuclei, and the coherence time of the electron spin (that is, the minimum hyperfine splitting required for detection of the nuclear spin). To investigate how many nuclei can be used as quantum register, we performed numerical simulations. The average number of suitable ^{13}C nuclear spins can be estimated by calculating the volume within the diamond of suitable nuclei positions. To this end, we assume purely magnetic dipole-dipole hyperfine interaction. The conditions for a suitable position are (1) that the hyperfine splitting is larger than a certain threshold value (for example, larger than the line broadening), and (2) that the difference of the nuclear spin eigenbasis $\{|n_i\rangle\}$ with respect to the z -basis $\{|z_i\rangle\}$ is smaller than $1 - |\langle n_i|z_i\rangle|^2 < 0.25 \times 10^{-4}$. Numerical integration over the volume surrounding the NV yields the effective volume of suitable ^{13}C positions. The average number of ^{13}C spins is calculated by the density of nuclei in the diamond

lattice and the ^{13}C concentration. Extended Data Fig. 2a shows the scaling of the average number of suitable ^{13}C spins per defect centre with the magnetic field B . As B is increased, the eigenbasis of the nuclear spins becomes mainly defined by the magnetic field, and independent of the electron spin state, which enables quantum-non-demolition readout. At some point, all available nuclear spins with large enough hyperfine splitting are suitable for single-shot readout, and the curves in Extended Data Fig. 2a saturate.

Depending on the hyperfine splitting and the electron spin coherence time, nuclear spins are divided into those strongly coupled and those weakly coupled. Strongly coupled means that the hyperfine splitting is larger than the electron spin inhomogeneous broadening ($1/T_2^*$), such that the nuclear spins can be directly addressed by the electron (compare Fig. 1b). The weak coupling case indicates that the hyperfine splitting is smaller than $1/T_2^*$, but larger than the electron spin homogeneous broadening ($1/T_2$). In this case, the nuclear spins can be addressed via the electron spin by dynamical decoupling¹³.

In this work, we focused on strongly coupled nuclear spins. For randomly distributed ^{13}C , the average number of suitable strongly coupled ^{13}C spins depending on B is shown in Extended Data Fig. 2a, for a ^{13}C concentration of 0.2% (bright red dots), as used here, and for the natural ^{13}C concentration of 1.1% (dark red dots). Additionally, this value could be increased by deliberate construction of the diamond surrounding the NV centre, for example, by implantation of molecules including ^{13}C atoms. This method would be limited by the addressability of all conditional electron spin transitions, for example transitions $m_S = 0 \leftrightarrow -1$. This means conditional on all possible nuclear spin states $|m_{1,1}, m_{1,2}, \dots, m_{1,N}\rangle$ of all N coupled nuclear spins, where the $m_{1,i}$ are the nuclear spin projections. For N nuclei, there are 2^N such transitions. The inhomogeneous linewidth of the NV electron spin transition can be small enough to resolve hyperfine interactions on the order of a few kHz (ref. 18). The maximum suitable hyperfine interaction of the closest ^{13}C spins is a few MHz. Therefore, in principle up to $\sim 10^3$ resonance lines are resolvable, resulting in ~ 10 addressable nuclear spins. Thus, the presented high-speed and high-fidelity CPhase gate is suitable for use with a maximum number of coupled nuclear spins of around 10. The above estimation, however, assumes that each of the mentioned coupled nuclear spins, if ordered by their coupling strength, increase their coupling strength by roughly a factor of two with respect to the previous nucleus.

The number of usable nuclear spins can be increased even further via addressing the nuclear spin resonance frequency space and by including weakly coupled nuclear spins. In that case, the number of transitions scales much more conveniently (that is, linearly with N); however, the fast CPhase gate is no longer practicable. For detection and manipulation of weakly coupled ^{13}C nuclear spins via the electron spin using dynamical decoupling¹³, the limiting factor is the homogeneous decoherence time T_2 of the electron spin. The T_2 time depends linearly on the concentration c of ^{13}C (ref. 32). For a diamond with natural abundance of ^{13}C (1.1%), a T_2 time of 0.65 ms has been reported³², corresponding to a minimal hyperfine interaction of ~ 1.5 kHz for a detectable ^{13}C . Note that for very low ^{13}C concentration, the T_2 time is not limited by the ^{13}C concentration, but by other effects³² (for example, longitudinal spin lattice relaxation, T_1). Extended Data Fig. 2a shows that in a diamond with natural abundance of ^{13}C (1.1%), on average ~ 10 suitable ^{13}C spins with hyperfine splitting of > 20 kHz are available per NV. Note that for smaller hyperfine splitting, the density of spins increases greatly (see Extended Data Fig. 2b), which makes selective addressing of these spins challenging.

Creation of an SIL. All measurements of the state of the NV and surrounding nuclear spins are based on detection of the NV fluorescence. Therefore, efficient photon detection is beneficial, as it increases measurement accuracy and reduces measurement time. For single-shot readout, measurement fidelity and hence state initialization fidelity is increased. However, much of the fluorescence of the NV is lost at the diamond surface due to total internal reflection. This problem can be overcome by creating an SIL³³. The SIL is a solid half-sphere with the emitter (NV) at its central point, such that the fluorescence of the NV will always hit the solid surface perpendicularly. Here, we created a SIL directly in the diamond around the NV by focused ion beam milling (Extended Data Fig. 3a). This increases the detected fluorescence of the NV by a factor of more than 3 (Extended Data Fig. 3b).

Details of the measurement sequence. The full measurement sequence in shown in Extended Data Fig. 4. The first step for every measurement is the initialization of nuclear spins. This is achieved by measuring the state of the nuclear spin register, until it is in the desired state. For technical reasons, we always perform the full measurement sequence, and finally only use the results for which the state was initialized. Because the probability for successful initialization decreases with the number of states, we also use swap-like gates to transfer the initialized electron spin state to each nuclear spin, as shown in Extended Data Fig. 4 and Fig. 2a. Owing to imperfect gates and non-perfect electron initialization, this increases the success probability of initialization of the whole register by a final single-shot measurement to around 50%. Another important issue is the charge state of the NV. Studies

have shown that for typical measurement conditions, the NV is in the undesired neutral charge state with around 30% probability, which shows short coherence times^{34,35}. Here, we use charge-state post-selection as demonstrated in ref. 36. To this end, a controlled π -pulse is applied on the electron spin only if the nuclear spins are in the initial state. Next, we apply a controlled π -pulse on the ^{14}N from $m_I = +1$ to 0, depending on the electron spin state. These two pulses will only work if the NV is properly initialized in its negative charge state. Otherwise, the ^{14}N will remain in the $m_I = +1$ state, which can be checked by a final single-shot measurement. After these two π -pulses, we start the actual experiment, as described in the main text, using the ^{14}N $m_I = 0, -1$ subspace as a qubit. Finally, we read out the state of the nuclear spins. First, we measure if the ^{14}N is still in the $m_I = +1$ state, in which case the NV was in the wrong charge state, and the result is discarded. Subsequently, we measure the state of the remaining qubit register by probing each state for separate runs of the experiment.

For the initialization of the nuclear spin register, we can in principle reach near-perfect initialization, by shifting the threshold in Fig. 1e further down. Extended Data Fig. 5 shows the measured initialization fidelity as a function of the duration of the single-shot readout (in number of readout steps) and the relative threshold shift. An initialization fidelity approaching 99% can be achieved. This method comes with a trade-off of increased measurement time, because the probability of a successful initialization will be reduced.

Optimal control of electron spin. Owing to the weak coupling to the two ^{13}C spins, selective rotations on specified transitions with standard rectangular microwave pulses will induce an additional unwanted operation on others that are off-resonance. To suppress the off-resonance effect and achieve a high-fidelity control of the electron spin, we apply optimal control to the electron spin. Two microwave frequencies (f_1, f_2) are imposed simultaneously to cover the whole spectrum and realize arbitrary control of the electron spin (Extended Data Fig. 6a). The amplitude and phase of each microwave frequency are taken as free parameters to obtain the desired control. The pulse sequence is optimized using the GRAPE algorithm^{28,37}. In the experiment, the optimal pulse sequence is realized by an arbitrary waveform generator (AWG; Agilent Technologies M8190A). Extended Data Fig. 6b shows an optimal pulse sequence applied to prepare a GHZ state of nuclear spins in the experiment. It is indeed a controlled phase (CPhase) gate on the electron spin, which maps the nuclear spin states $|100\rangle$ and $|111\rangle$ to states $-|100\rangle$ and $-|111\rangle$, respectively. The calculated fidelity is above 99% even for small detuning (± 20 kHz), which means that the pulse is robust in the presence of inhomogeneous line broadening of the electron spin. In the actual implementation, the instability of the microwave amplitude is found to be one factor limiting the performance of the pulse.

Mermin inequality. A first test for the non-deterministic and non-local structure of quantum mechanics was proposed by Bell in 1964 for the case of two entangled particles³⁸. A general inequality to measure this behaviour of quantum mechanics for multiple particles was derived by Mermin³⁰. For three particles, the inequality is

$$|\langle \sigma_x \sigma_x \sigma_y \rangle + \langle \sigma_x \sigma_y \sigma_x \rangle + \langle \sigma_y \sigma_x \sigma_x \rangle - \langle \sigma_y \sigma_y \sigma_y \rangle| \leq 2 \quad (3)$$

for any deterministic, local theory. However, the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|000\rangle + i|111\rangle) \quad (4)$$

yields the value 4 for inequality (3). In our case, the prepared state is

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\bar{y}\bar{y}\rangle + i|1yy\rangle) \quad (5)$$

where $|y\rangle = |0\rangle + i|1\rangle$ and $|\bar{y}\rangle = |0\rangle - i|1\rangle$. This state can be derived from equation (4) by $\pi/2$ rotations around the x axis on qubits two and three (second and third value in the state description). Thereby, σ_y in inequality (3) becomes σ_z for these two qubits, yielding the inequality:

$$|\langle \sigma_x \sigma_x \sigma_z \rangle + \langle \sigma_x \sigma_z \sigma_x \rangle + \langle \sigma_y \sigma_x \sigma_x \rangle - \langle \sigma_y \sigma_z \sigma_z \rangle| \leq 2 \quad (6)$$

To measure the single terms, we transform each term to $\langle I\sigma_z \rangle$ (where I is the identity). This can be achieved by a unitary operation U according to $\langle \psi|A|\psi\rangle = \langle \psi'|A'|\psi'\rangle$ for $|\psi'\rangle = U|\psi\rangle$ and $A' = UAU^\dagger$. The operations are shown in Extended Data Table 1. The Cphase gate is described in the main text; here it is conditional on the states $|011\rangle$ and $|101\rangle$.

State tomography. The density matrix of the entangled GHZ and W state is measured by state tomography. Therefore, the density matrix ρ is expanded as

$$\rho = \sum_i a_i A_i \quad (7)$$

where the A_i form a basis of mutually orthogonal operators. In the case of a single qubit, we would use the Pauli matrices $I, \sigma_x, \sigma_y, \sigma_z$. For multiple qubits, the A_i are all combinations of the Pauli matrices for all qubits, that is, $A_i = \{I^{(1)}I^{(2)} \dots I^{(n)}, I^{(1)}I^{(2)} \dots \sigma_x^{(n)}, \dots, \sigma_z^{(1)}\sigma_z^{(2)} \dots \sigma_z^{(n)}\}$, where the superscripts indicate on which qubit the operator acts. The a_i in equation (7) are given by the expectation value $\text{Tr}(A_i\rho)$ of the corresponding operator. These expectation values can be determined experimentally. Because only the expectation value of the σ_z operator can be measured directly, we apply $\pi/2$ rotations around the y and x axes of a spin to get the expectation values of σ_x and σ_y . Solving equation (7) with the measured a_i yields the state of the system. Note that owing to statistical measurement errors, the measured density matrix may not be positive semi-definite.

Process tomography. A general characterization of a quantum mechanical process \mathcal{E} on a system is obtained by expanding the density matrix ρ into a basis of orthonormal states $|\psi_i\rangle$, and measuring the effect of the operator E_i of the process on each of these states:

$$\begin{aligned} \rho &= \sum_i p_i |\psi_i\rangle\langle\psi_i|, \\ \mathcal{E}(\rho) &= \sum_i E_i p_i |\psi_i\rangle\langle\psi_i| E_i^\dagger \end{aligned} \quad (8)$$

The E_i can be measured by initializing the states $|\psi_i\rangle$ and performing state tomography after the process. Furthermore, we can do the expansion

$$E_i = \sum_m \chi_{im} A_m \quad (9)$$

where the A_m form a basis of mutually orthogonal operators. Inserting equation (9) into equation (8) yields

$$\mathcal{E}(\rho) = \sum_{mn} \chi_{mn} A_m \rho A_n^\dagger \quad (10)$$

where the $\chi_{mn} = \sum_i \chi_{im} \chi_{in}$ form the process matrix χ .

For the error correction, we only consider the effect of the process on the qubit that is carrying the information after the correction. Additionally, we are only interested in the fidelity F of the measured process and the ideal process χ_{id} , which is:

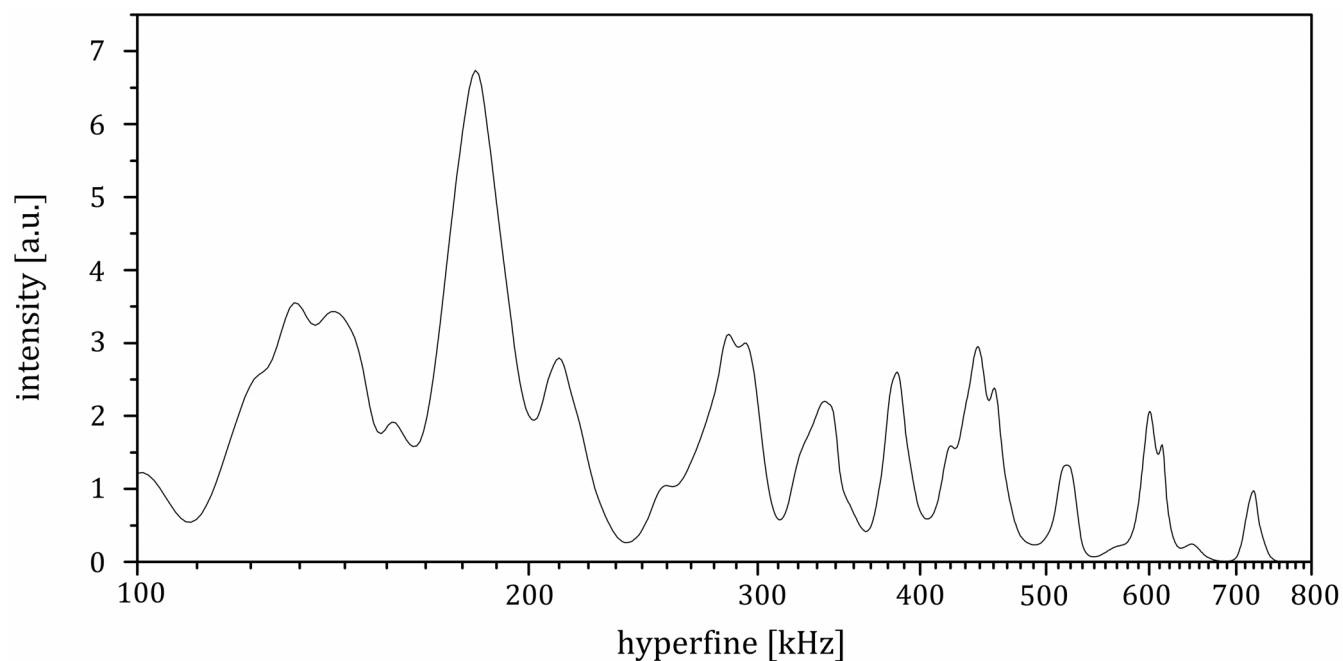
$$F = \text{Tr}(\chi_{id}\chi) \quad (11)$$

Here, we chose the Pauli matrices $A_m = \{I, \sigma_x, \sigma_y, \sigma_z\}$ as the operator basis. The ideal process for the error correction is the identity, I , and therefore $\chi_{11}^{id} = 1$ with all other values being zero. The fidelity is then χ_{11} . Extended Data Table 2 shows the measurements we perform to obtain

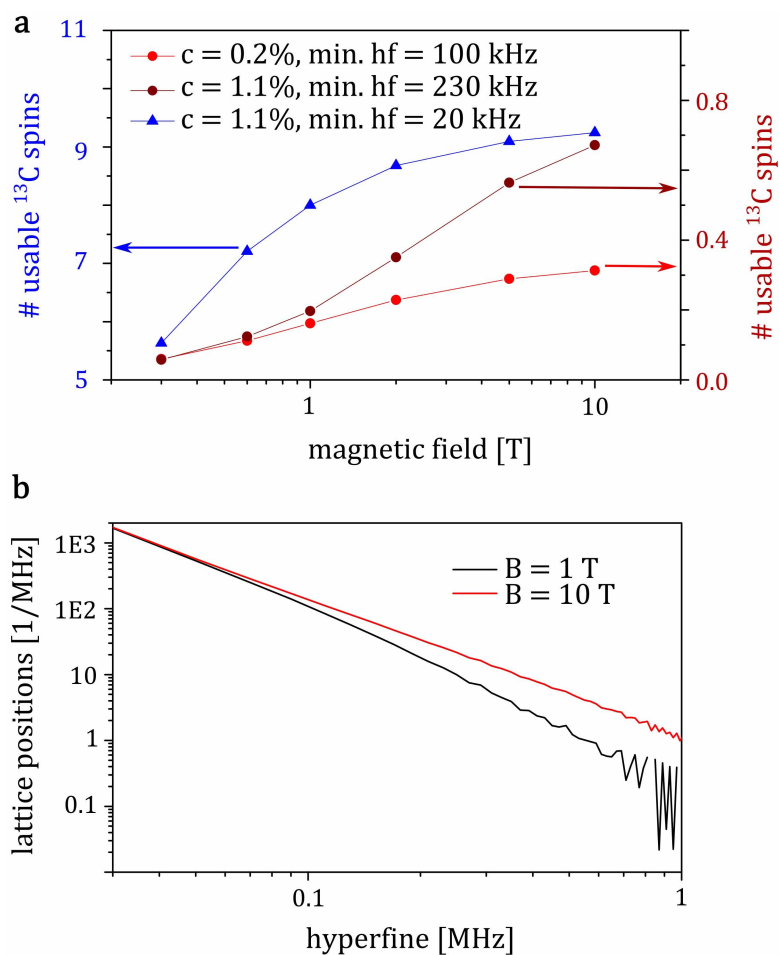
$$\chi_{11} = \frac{1 + (r_{z,z} - r_{-z,z} + r_{x,x} - r_{-x,x} + r_{y,y} - r_{-y,y})}{4} \quad (12)$$

by noting that $\chi_{11} + \chi_{22} + \chi_{33} + \chi_{44} = 1$.

32. Mizuochi, N. *et al.* Coherence of single spins coupled to a nuclear spin bath of varying density. *Phys. Rev. B* **80**, 041201 (2009).
33. Marseglia, L. *et al.* Nanofabricated solid immersion lenses registered to single emitters in diamond. *Appl. Phys. Lett.* **98**, 133107 (2011).
34. Waldherr, G. *et al.* Dark states of single nitrogen-vacancy centers in diamond unraveled by single shot NMR. *Phys. Rev. Lett.* **106**, 157601 (2011).
35. Aslam, N., Waldherr, G., Neumann, P., Jelezko, F. & Wrachtrup, J. Photo-induced ionization dynamics of the nitrogen vacancy defect in diamond investigated by single-shot charge state detection. *New J. Phys.* **15**, 013064 (2013).
36. Waldherr, G. *et al.* High dynamic range magnetometry with a single nuclear spin in diamond. *Nature Nanotechnol.* **7**, 105–108 (2011).
37. Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbruggen, T. & Glaser, S. J. Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172**, 296–305 (2005).
38. Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).
39. Waldherr, G., Neumann, P., Huelga, S., Jelezko, F. & Wrachtrup, J. Violation of a temporal Bell inequality for single spins in a diamond defect center. *Phys. Rev. Lett.* **107**, 090401 (2011).



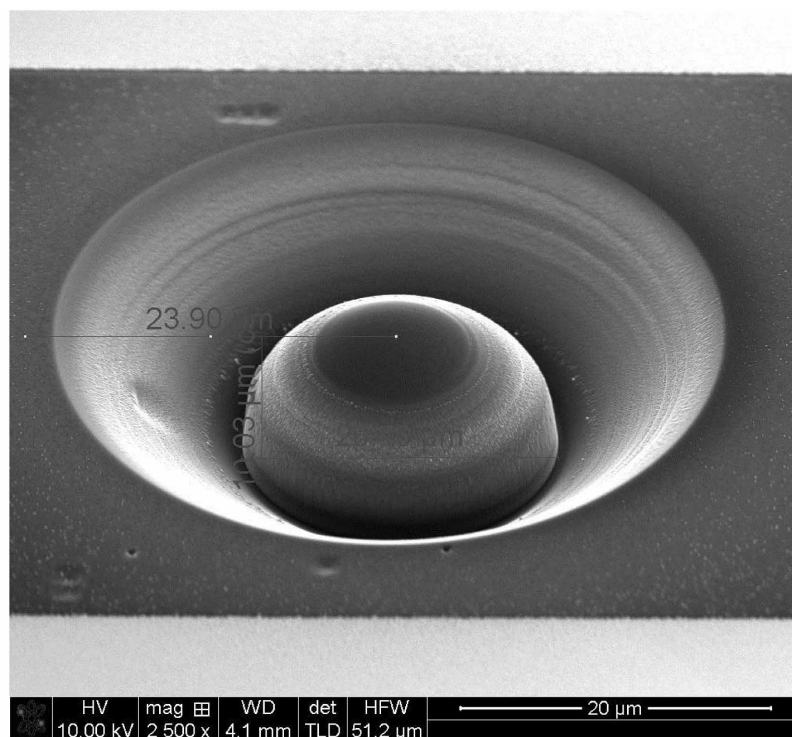
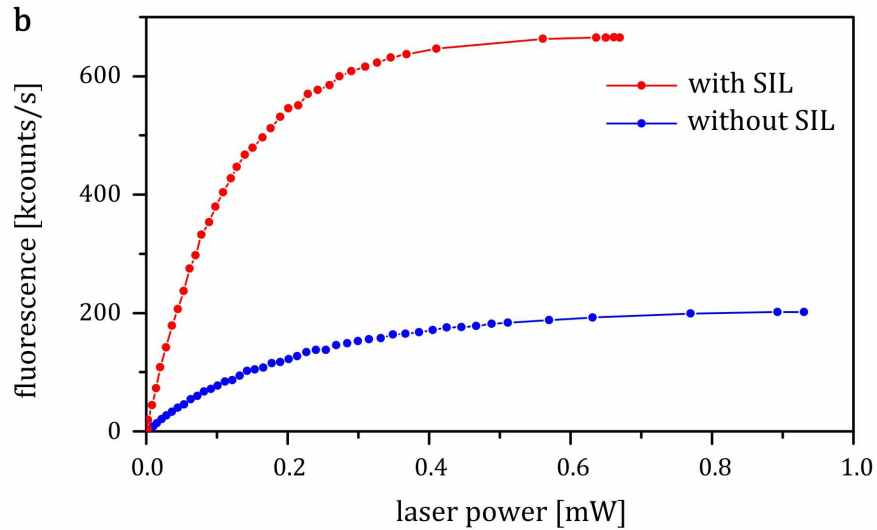
Extended Data Figure 1 | Hyperfine spectrum as explained in the main text. Only hyperfine values with an error of less than 4% were used, such that hyperfine values which are close to each other can be resolved.



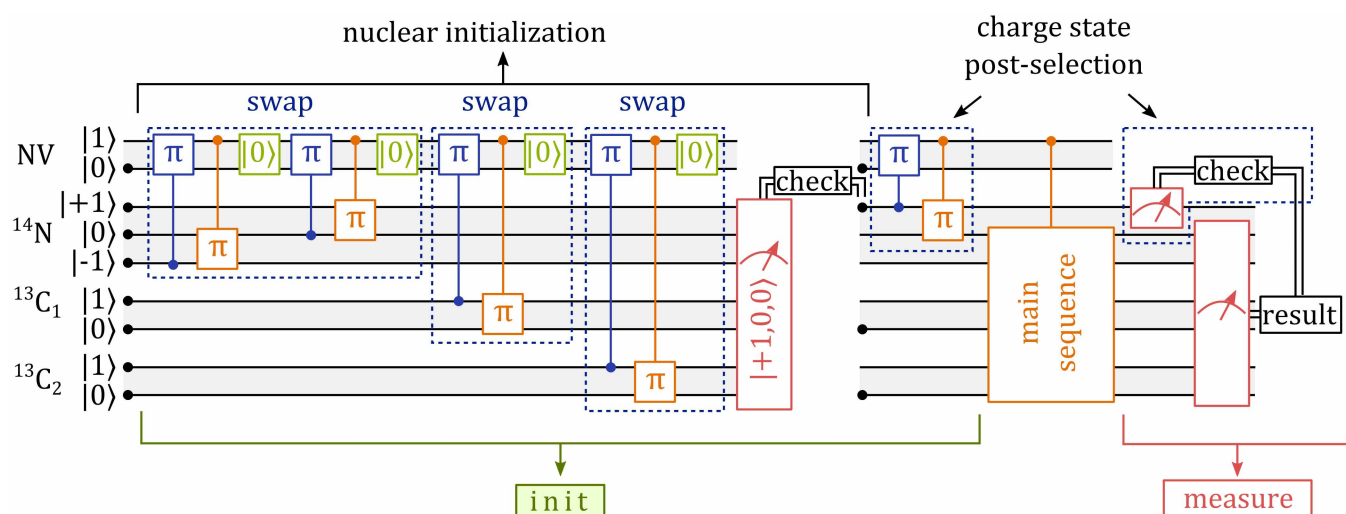
Extended Data Figure 2 | Estimation of number of usable ^{13}C spins.

a, Average number of suitable ^{13}C spins per NV for different ^{13}C concentrations c and different minimum hyperfine (min. hf.) interactions. Red dots, strongly coupled nuclei; blue triangles, effect of including weakly coupled nuclei with at least 20 kHz hyperfine splitting. **b**, Spectral density of suitable lattice

positions per NV for different magnetic fields B . Note that for these simulations, actual lattice positions are not taken into account. The fluctuations at higher hyperfine interaction are due to numerical grain, that is, due to discretization of the integration volume.

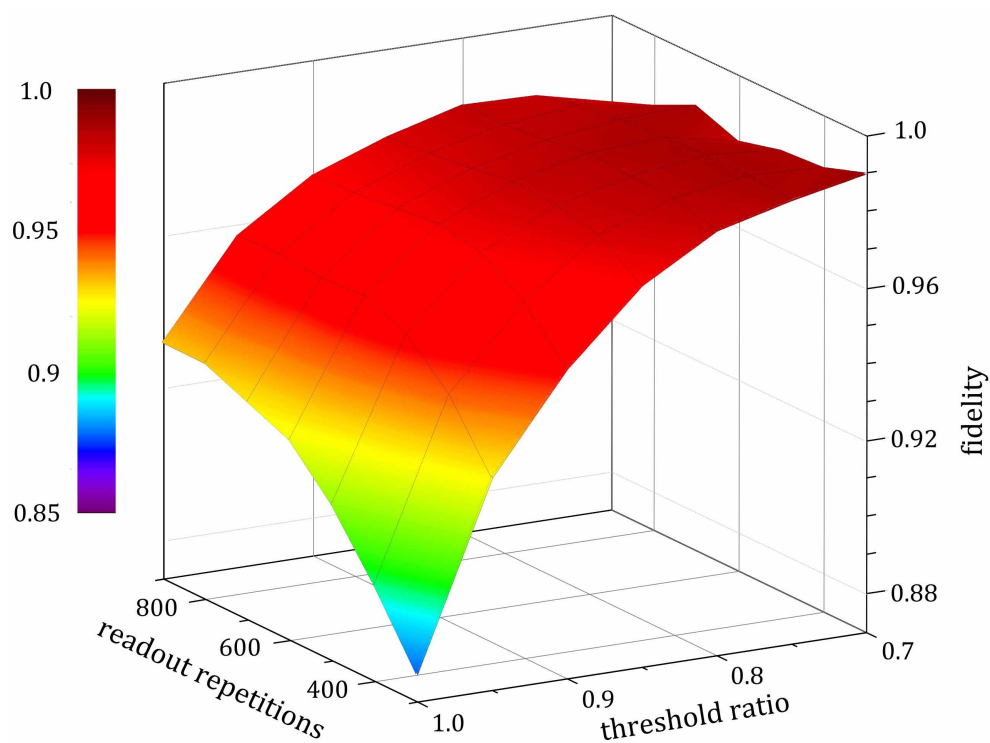
a**b**

Extended Data Figure 3 | Solid immersion lens (SIL). **a**, Image of the SIL in diamond. **b**, Saturation curves of the NV with and without the SIL (measurements were performed with a oil-immersion objective).



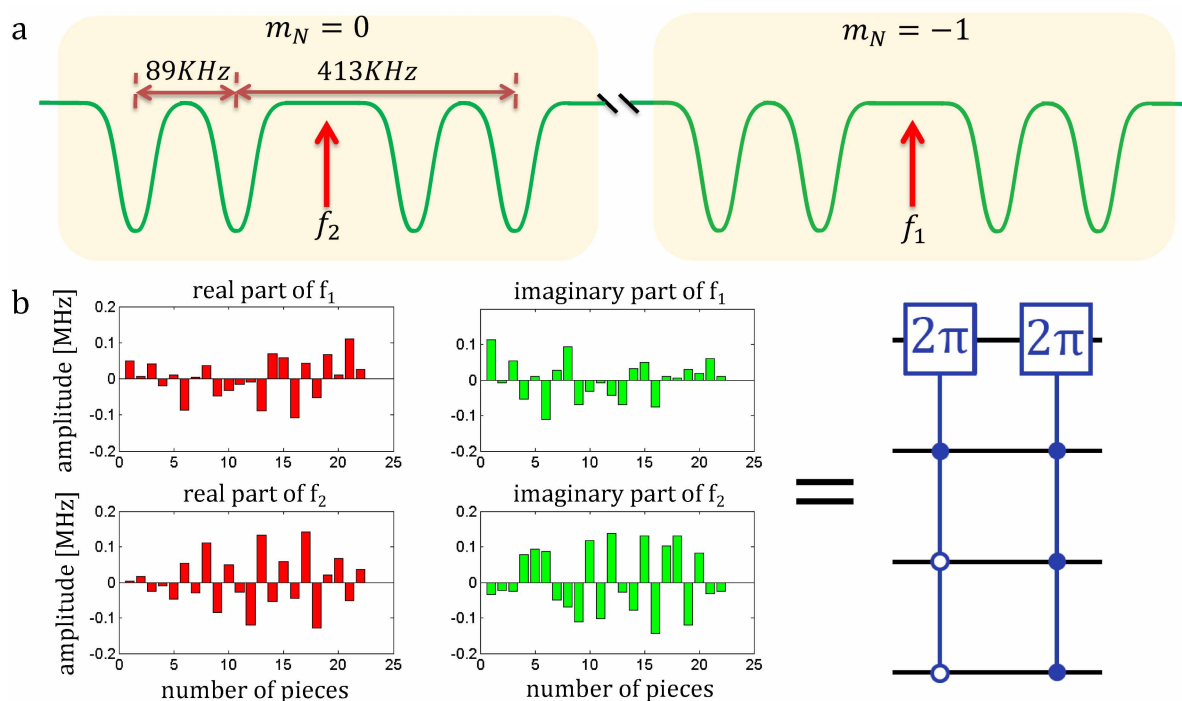
Extended Data Figure 4 | Full sequence for initialization and readout of the nuclear register. The 'main sequence' part is the actual quantum algorithm,

including final local rotations for setting the measurement basis. Note that the charge state post-selection can be substituted by charge state pre-selection^{34,39}.



Extended Data Figure 5 | Initialization fidelity of the whole spin register as a function of the number of readout repetitions and the relative shift of

the initialization threshold. The fidelity is colour-coded according to the colour bar.



Extended Data Figure 6 | Optimal control. **a**, The two microwave frequencies f_1 and f_2 , relative to the electron spin transition frequencies, applied in the experiment. **b**, The pulse sequence on the left side shows the

piecewise-constant control amplitudes (Rabi frequency) for the real and imaginary parts of f_1 and f_2 , where each piece (bar) has a duration of 1.46 μ s. It realizes the controlled gate on the electron spin given on the right side.

Extended Data Table 1 | Measurement procedure for the Mermin inequality

Measure	Operation
$\langle \sigma_x \sigma_x \sigma_z \rangle$	$\left(\frac{\pi}{2}\right)_{1,y} \left(\frac{\pi}{2}\right)_{2,y} \left(-\frac{\pi}{2}\right)_{3,y} \text{ (CPhase) } \left(\frac{\pi}{2}\right)_{3,y}$
$\langle \sigma_x \sigma_z \sigma_x \rangle$	$\left(\frac{\pi}{2}\right)_{1,y} \text{ (CPhase) } \left(\frac{\pi}{2}\right)_{3,y}$
$\langle \sigma_y \sigma_x \sigma_x \rangle$	$\left(\frac{\pi}{2}\right)_{1,x} \left(\frac{\pi}{2}\right)_{2,y} \text{ (CPhase) } \left(\frac{\pi}{2}\right)_{3,y}$
$\langle \sigma_y \sigma_z \sigma_z \rangle$	$\left(\frac{\pi}{2}\right)_{1,x} \left(-\frac{\pi}{2}\right)_{3,y} \text{ (CPhase) } \left(\frac{\pi}{2}\right)_{3,y}$

Extended Data Table 2 | Measurement procedure and theoretical results for the process fidelity

Init	Measure	result
$ z\rangle$	σ_z	$r_{z,z} = \chi_{11} + \chi_{14} + \chi_{41} + \chi_{44}$
$ -z\rangle$	σ_z	$r_{-z,z} = -\chi_{11} + \chi_{14} + \chi_{41} - \chi_{44}$
$ x\rangle$	σ_x	$r_{x,x} = \chi_{11} + \chi_{12} + \chi_{21} + \chi_{22}$
$ -x\rangle$	σ_x	$r_{-x,x} = -\chi_{11} + \chi_{12} + \chi_{21} - \chi_{22}$
$ y\rangle$	σ_y	$r_{y,y} = \chi_{11} + \chi_{13} + \chi_{31} + \chi_{33}$
$ -y\rangle$	σ_y	$r_{-y,y} = -\chi_{11} + \chi_{13} + \chi_{31} - \chi_{33}$

Mosaic two-lengthscale quasicrystals

T. Dotera¹, T. Oshiro¹ & P. Ziherl^{2,3}

Over the past decade, quasicrystalline order¹ has been observed in many soft-matter systems: in dendritic micelles², in star³ and tetrablock⁴ terpolymer melts and in diblock copolymer⁵ and surfactant micelles⁶. The formation of quasicrystals^{7–9} from such a broad range of ‘soft’ macromolecular micelles suggests that they assemble by a generic mechanism rather than being dependent on the specific chemistry of each system. Indeed, micellar softness has been postulated⁷ and shown to lead to quasicrystalline order¹⁰. Here we theoretically explore this link by studying two-dimensional hard disks decorated with step-like square-shoulder repulsion that mimics, for example, the soft alkyl shell around the aromatic core in dendritic micelles². We find a family of quasicrystals with 10-, 12-, 18- and 24-fold bond orientational order which originate from mosaics of equilateral and isosceles triangles formed by particles arranged core-to-core and shoulder-to-shoulder. The pair interaction responsible for these phases highlights the role of local packing geometry in generating quasicrystallinity in soft matter, complementing the principles that lead to quasicrystal formation in hard tetrahedra^{11,12}. Based on simple interparticle potentials, quasicrystalline mosaics may well find use in diverse applications ranging from improved image reproduction¹³ to advanced photonic materials¹⁴.

The simple rules needed to encode the quasicrystalline order are as remarkable as the order itself. All that is required to create a one-dimensional quasicrystal is the Fibonacci sequence of long and short segments such that their length and number ratios are equal to the golden ratio $\tau = (1 + \sqrt{5})/2 \approx 1.618$. Also pervaded by τ is the archetypal Penrose tiling, the two-dimensional analogue of the Fibonacci quasicrystal composed of a small set of prototiles, such as the fat and the skinny Penrose rhombi¹⁵. Because the golden ratio represents the geometry of the regular pentagon, Penrose tiling must have a non-crystallographic tenfold symmetry directly related to the shape of the Robinson triangles forming the prototiles¹⁵.

The Robinson triangles are a very useful concept for constructing quasicrystals from particles because in two dimensions the local structure of three particles with a two-lengthscale pair interaction is either an isosceles or an equilateral triangle (Fig. 1c). Indeed, properly tailored shoulder-like interactions as well as single-well and double-well interactions (examples are simple step-like potentials¹⁶ and hard-core repulsion combined with square-well attraction¹⁷) do induce the formation of decagonal^{17–19} and dodecagonal quasicrystals^{16,19,20}. Here we propose a bottom-up framework for two-lengthscale quasicrystals that is analogous to the canonical-cell tiling model²¹, showing that such quasicrystals

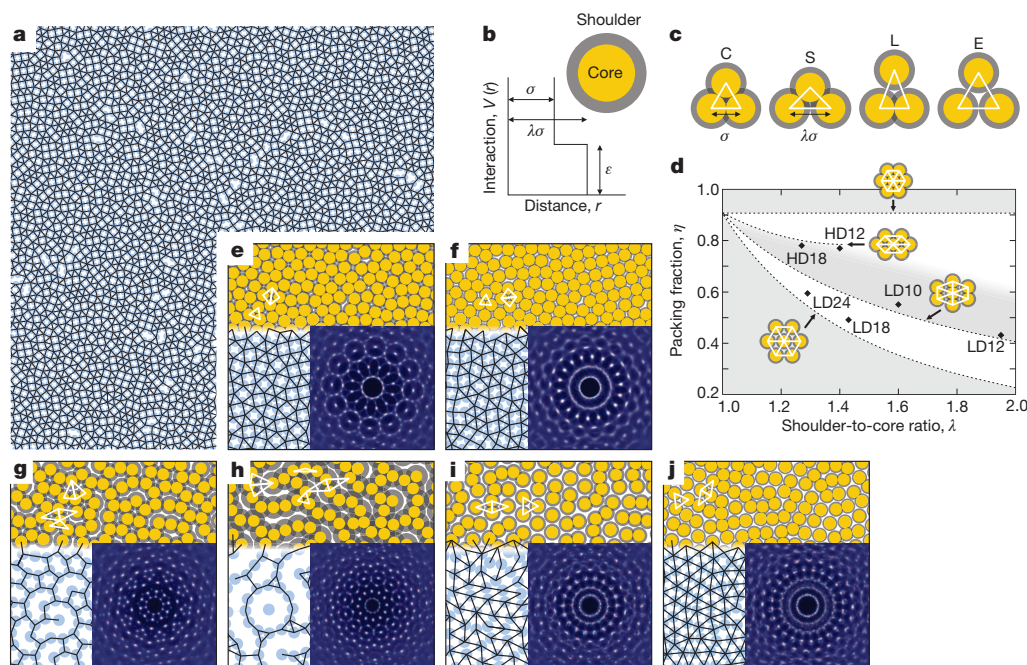


Figure 1 | Real-space structures and diffraction patterns. **a**, Snapshot illustrating the mosaic nature of two-lengthscale quasicrystals; shown here is the core-only representation of the HD12 phase with shoulder-to-core ratio $\lambda = 1.40$, reduced temperature $\Theta = k_B T/\epsilon = 0.278$ and packing fraction $\eta = \pi N \sigma^2/4A = 0.770$ where N is the number of disks in a box of area A . **b**, Hard-core/square-shoulder pair interaction. **c**, Local disk packings: collapsed equilateral (C), short isosceles (S), long isosceles (L) and expanded equilateral triangle (E). **d**, Location of quasicrystals in the phase diagram. Dashed lines

represent two hexagonal phases built from C and E triangles and two rhombic phases built from S and L triangles, respectively. **e–j**, HD12 (**e**; same parameters as in **a**), HD18 (**f**; $\lambda = 1.27$, $\Theta = 0.208$, $\eta = 0.780$), LD10 (**g**; $\lambda = 1.60$, $\Theta = 0.133$, $\eta = 0.550$) and LD12 (**h**; $\lambda = 1.95$, $\Theta = 0.154$, $\eta = 0.430$), LD18 (**i**; $\lambda = 1.43$, $\Theta = 0.0885$, $\eta = 0.490$) and LD24 (**j**; $\lambda = 1.29$, $\Theta = 0.098$, $\eta = 0.595$) phases. Bonds drawn in the core-only representation (**a** and the bottom-left parts of panels **e–j**) accentuate the polygonal tiles.

¹Department of Physics, Kinki University, 3-4-1 Kowakae, Higashi-Osaka 577-8502, Japan. ²Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia. ³Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia.

include phases of 18- and 24-fold symmetry in addition to the previously known 10- and 12-fold quasicrystals. Motivated by the hard-sphere theory of fluid-to-solid transition²², we numerically explored two-dimensional disks interacting with the purely repulsive hard-core/square-shoulder pair potential^{23,24} (Fig. 1b):

$$V(r) = \begin{cases} \infty, & r < \sigma \\ \varepsilon, & \sigma < r < \lambda\sigma \\ 0, & r > \lambda\sigma \end{cases}$$

The shoulder height ε sets the energy scale, and the shape of the potential is described solely by the ratio of the shoulder diameter $\lambda\sigma$ to the hard-core diameter σ .

Representative real-space images of the quasicrystals found are shown in Fig. 1 along with Fourier transforms of particle positions. At a shoulder-to-core ratio of $\lambda \approx 1.40$, we find a well-ordered 12-fold high-density (HD12) quasicrystal (Fig. 1a, e) recently reported in attractive particles interacting with a double-well potential of a similar lengthscale ratio²⁵. The most interesting feature of the HD12 phase is the value of λ , which is considerably smaller than $\lambda = \sqrt{2} + \sqrt{3} \approx 1.932$, at which dodecagonal phases were observed in related systems^{16,19,20}. The HD12 phase can be regarded as a random square-triangle tiling²⁶, its triangle-to-square ratio being close to the ideal value of $4/\sqrt{3} \approx 2.309$ (ref. 27). At $\lambda \approx 1.27$, we find a high-density phase with 18-fold symmetry (Fig. 1f). With a large fraction of particles forming close-packed equilateral triangles (C triangle in Fig. 1c), the HD18 quasicrystal is structurally similar to the HD12 phase except that it contains rhombi of two 80° short isosceles triangles (S triangle in Fig. 1c) rather than squares.

Also new is the low-density 12-fold phase (LD12, Fig. 1h) based on C triangles and skinny rhombi formed by long isosceles triangles (L triangle in Fig. 1c). Its main structural features are regular dodecagons as well as straight, winding and meandering stripes reminiscent of labyrinthine phases²⁴. At $\lambda \approx 1.60$, we observe the decagonal LD10 phase (Fig. 1g) constructed from sequences of the Robinson L and S triangles, which form a standard set of decagonal tiles including pentagons and dodecagons¹⁹. The dominant structural features of the ill-ordered LD18 and LD24 phases (Fig. 1i, j) are skinny rhombi of two L triangles mixed with the shoulder-to-shoulder equilateral E triangles (see Fig. 1c). The location of quasicrystals in the phase diagram relative to the two hexagonal and rhombic lattices is shown in Fig. 1d.

As illustrated by the snapshot of the HD12 phase in Fig. 1a, as well as by Supplementary Figs 3, 6, 10, 14, 17 and 20, our quasicrystals may be viewed as mosaics of small crystallites packed such that the quasicrystalline orientational order is established on a lengthscale considerably larger than the nearest-neighbour distance. In this respect, they are quite different from conventional metallic-alloy quasicrystals with a local structure based on icosahedral, decagonal or dodecagonal clusters of atoms. The mosaic nature of quasicrystals presented in Fig. 1 is consistent with the short range of particle-particle interactions, which do not extend beyond nearest neighbours in either the high- or low-density phases.

To further quantify our quasicrystals, we computed the bond-orientational-order correlation functions $G_m(r)$ for their respective dominant modes m (for example, $m = 18$ in HD18 and LD18 phases), where r is the distance between two bonds. Figure 2 shows that the correlation function in the more classical LD10 levels off at a constant value, indicating long-range orientational order and implying that this phase may be referred to as random decagonal. Long-range order is also present in HD12 and LD12 phases, so that both are true random dodecagonal quasicrystals. On the other hand, correlations in the 18- and the 24-fold quasicrystals exhibit a power-law decay of an exponent which varies across the phases as well as with temperature. In analogy with the hexatic phase (Supplementary Discussion I), these quasicrystals are best described as the decaoctatic and the icositettratic phase, respectively.

An interesting consequence of the rather small shoulder-to-core ratio in the square-triangle HD12 phase concerns its thermodynamic stability. A repulsive shoulder with λ just a little smaller than $\sqrt{2}$ reaches only to

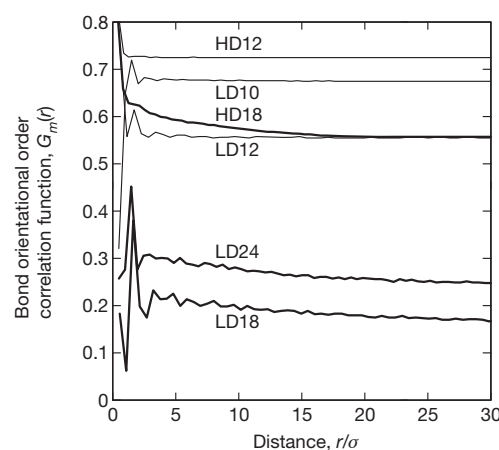


Figure 2 | Bond orientational order. The 10-, 12-, 18- and 24-fold bond orientational order correlation functions characteristic of the respective quasicrystals shown in Fig. 1; r/σ is the distance expressed in units of hard-core diameter σ .

the nearest neighbours and so the energy of the tiling depends on the relative number of squares and triangles but not on the structure of the tiling, implying that all defect-free tilings have the same energy as required in the maximally random tiling²⁶ (Supplementary Discussion V). Thus at a density for which the triangle-to-square ratio is ideal, the minimal-energy state is a disordered dodecagonal quasicrystal rather than a crystal, meaning that the HD12 phase may be stable both at intermediate and low temperatures.

Additional insight into the stability of the HD12, LD12, LD10, and HD18 quasicrystals is provided by the rapid growth of bond order in the initial random state (Supplementary Discussion IIA–D). This suggests that these phases are preferred over disordered (HD12 and HD18) and fluid (LD12 and LD10) states, so that at intermediate temperatures they must be (at least) very long-lived metastable states, much like the dodecagonal quasicrystals based on hard tetrahedra¹¹. The LD18 and LD24 phases form less readily and contain many defects (Supplementary Discussion IIE and IIF), calling for a more refined search in the temperature–density– λ phase space.

These new quasicrystal phases can be interpreted by generalizing the well-known random LD10²⁸ and HD12 quasicrystals²⁶ based on the hierarchy where two or more types of the canonical triangles shown in Fig. 1c are used to construct a set of equilateral polygons that tile the plane (Fig. 3a and Supplementary Fig. 26). Derived from the Robinson triangles, the decagonal LD10 equilateral polygons include the pentagon, hexagon, nonagon, U-tile and decagon^{19,28} as well as their combinations (Supplementary Fig. 24). On the other hand, the dodecagonal HD12 equilateral polygons comprise the C triangle and the square built from two S triangles with $\lambda = \sqrt{2}$ (Supplementary Fig. 25).

This hierarchy is also applicable at certain other shoulder-to-core ratios, the allowed values being determined by conditions which can be illustrated by considering the rhombus-triangle tiling. Unlike squares, two rhombi can be stacked either such that their short diagonals are parallel (as in the 4_a^4 cluster in Fig. 3b) or in a zigzag fashion (as in the 4_b^4 cluster in Fig. 3b). To admit the zigzag rhombus sequence in a random tiling, two nontrivial local tile clusters are needed—one containing a pair of adjacent rhombi with their acute angles pointing to the vertex and the other containing a pair of rhombi in the obtuse orientation. In the HD18 phase, there exist three clusters that accomplish this, all including three rhombi and one or two triangles (Fig. 3b). We refer to these clusters as special because they are only compatible with the 80°–100° shapes of rhombi. Also present are five generic clusters containing either zero, three or six triangles and four, two and zero rhombi, respectively.

The LD12 phase is composed of polygonal tiles including the C triangle, the regular dodecagon, and several winding and meandering shapes

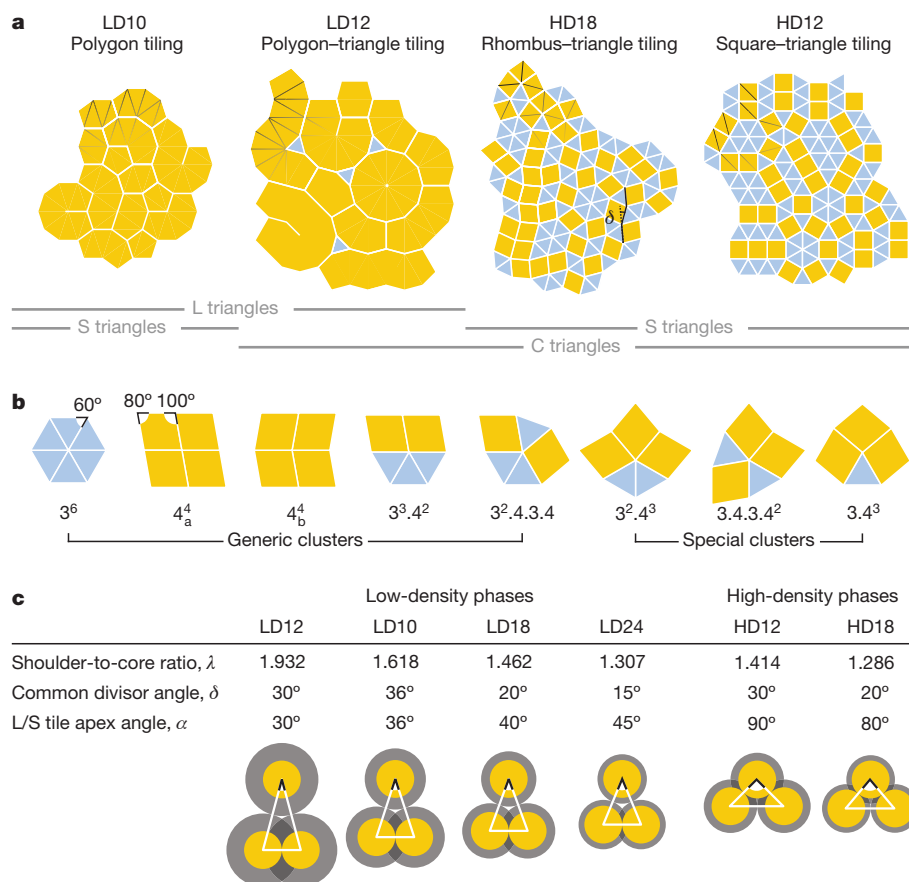


Figure 3 | Two-lengthscale triangle tilings. **a**, Comparison of structural features of the LD10, LD12, HD18 and HD12 tilings. The contours of the L and S triangles are shown in the top-left corners, and the black wiggly line in HD18 quasicrystal illustrates the angular mismatch δ . **b**, Generic and special HD18 vertices. **c**, Parameters of the six quasicrystals formed by

hard-core/square-shoulder disks: the optimal shoulder-to-core ratio, the greatest common divisor of the apex angles of the equilateral C and E tiles and the isosceles L and S triangles, and the apex angle (indicated in black on the diagrams below) of the L and S triangles.

(Fig. 3a). These tiles are based on parallelograms formed by L triangles of apex angle of 30° rather than on rhombi, and the special clusters combine the parallelograms and the C triangles (Supplementary Figs 32–34). Also determined by the special-cluster requirement is the shape of 40°–140° and 45°–135° rhombi. The LD18 and LD24 phases do contain special clusters with two or more rhombi in the acute orientation but lack an analogue of the 3.4³ special cluster of the HD18 phase with two or more rhombi in the obtuse orientation. Its absence is compensated by approximate special clusters consisting of S triangles as well as L and E tiles (Supplementary Figs 35 and 36). In the approximate special clusters the sum of angles at the vertex departs slightly from 360°, leading to a local distortion of the ideal two-lengthscale geometry. This is a plausible origin for the ill-ordered structure of LD18 and LD24 phases.

The symmetry of the HD12, HD18, LD18 and LD24 rhombus-triangle quasicrystals is determined by the rhombus acute angle α . As illustrated by the wiggly line in the HD18 tiling in Fig. 3a, combining equilateral triangles with rhombi introduces a set of bond directions at an angle of $\delta = |\alpha - 60^\circ|$ relative to the six directions of the hexagonal lattice. The angular mismatch δ is the greatest common divisor of α and 60° and in the HD18 phase shown in Fig. 3a, δ is 20°, giving a total of 18 bond orientations. Given that all greatest common divisors are unit fractions of 60° (Fig. 3c), our quasicrystals are characterized by $6n$ -fold orientational order, where $n = 2, 3$ and 4. On the other hand, the symmetry of the LD10 and LD12 phases is governed by the shape of the L triangle alone.

The previously unconsidered rhombus-triangle model of the 18-fold symmetry offers a simple interpretation of the temperature-induced transition from the face-centred cubic crystal to the 12-fold quasicrystal and then to the 18-fold quasicrystal reported in polyisoprene-poly

(ethylene oxide) block-copolymer micelles⁵. Within our model, two-dimensional variants of these structures are located very close to each other in the phase diagram (Fig. 1d) and a possible scenario for the transition could rely on the well-known temperature-dependent hydration of poly(ethylene oxide) chains²⁹, which constitute the hydrophilic shell of the micelles. Indeed, the hexagonal-HD12-to-HD18 transition is obtained by cooling the ensemble from a reduced temperature $\Theta = k_B T / \epsilon = 1.1$ to 0.14 accompanied by a moderate 10% decrease of the shoulder-to-core ratio from $\lambda = 1.40$ to 1.27 (Supplementary Discussion IV and Supplementary Video).

Our findings provide a conceptual understanding of the soft quasicrystals^{2–6} seen in deformable macromolecular particles primarily interacting with each other upon contact rather than at a distance. Such a microscopic interaction is perfectly consistent with our quasicrystals, which are stabilized almost exclusively by nearest-neighbour repulsion. We expect that the novel quasicrystals predicted within the present framework will be observed in core-shell soft spheres, including block copolymer micelles, microgels, nanogels and polymer-grafted organic and inorganic nanoparticles in which the relative sizes of core and shell can be controlled by temperature, pH, ionic strength and other stimuli³⁰.

METHODS SUMMARY

We used Monte Carlo simulations of the NVT ensemble of $N = 4,000$ particles in a square box with periodic boundary conditions (see Supplementary Discussion III). A typical cycle consisted of $2\text{--}8 \times 10^8$ equilibration steps in all quasicrystals except the LD18 and LD24 phases, which required a longer equilibration of about 3×10^9 steps. Equilibration was followed by about 10^7 sampling steps where several structural quantities, including bond orientational order parameters, were evaluated. The

simulated diffraction patterns are Fourier transforms of the centres of mass of the particles.

The bond orientational correlation function is defined by $G_m(r) = \left| n_r^{-1} \sum_{|\mathbf{r}_1 - \mathbf{r}_2| = r} \exp[i m(\theta_{\mathbf{r}_1} - \theta_{\mathbf{r}_2})] \right|$, where m is an integer characterizing the symmetry of the bond-orientational order, the sum is over all bonds separated by r and n_r is their number; $\theta_{\mathbf{r}_1}$ and $\theta_{\mathbf{r}_2}$ are the angles between an arbitrary axis and bonds connecting particles separated by no more than the shoulder diameter $\lambda\sigma$ and located at \mathbf{r}_1 and \mathbf{r}_2 , respectively. The values of $G_m(r)$ shown in Fig. 2 were obtained by averaging over 20 samples.

Received 7 June; accepted 2 December 2013.

Published online 2 February 2014.

- Levine, D. & Steinhardt, P. J. Quasicrystals—a new class of ordered structures. *Phys. Rev. Lett.* **53**, 2477–2480 (1984).
- Zeng, X. *et al.* Supramolecular dendritic liquid quasicrystals. *Nature* **428**, 157–160 (2004).
- Hayashida, K., Dotera, T., Takano, A. & Matsushita, Y. Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* **98**, 195502 (2007).
- Zhang, J. & Bates, F. S. Dodecagonal quasicrystalline morphology in a poly(styrene-*b*-isoprene-*b*-styrene-*b*-ethylene oxide) tetrablock terpolymer. *J. Am. Chem. Soc.* **134**, 7636–7639 (2012).
- Fischer, S. *et al.* Colloidal quasicrystals with 12-fold and 18-fold diffraction symmetry. *Proc. Natl Acad. Sci. USA* **108**, 1810–1814 (2011).
- Xiao, C., Fujita, N., Miyasaka, K., Sakamoto, Y. & Terasaki, O. Dodecagonal tiling in mesoporous silica. *Nature* **487**, 349–353 (2012).
- Ungar, G. & Zeng, X. Frank-Kasper, quasicrystalline and related phases in liquid crystals. *Soft Matter* **1**, 95–106 (2005).
- Lifshitz, R. & Diamant, H. Soft quasicrystals—why are they stable? *Phil. Mag.* **87**, 3021–3030 (2007).
- Mikhael, J., Roth, J., Helden, L. & Bechinger, C. Archimedean-like tiling on decagonal quasicrystalline surfaces. *Nature* **454**, 501–504 (2008).
- Iacovella, C. R., Keys, A. S. & Glotzer, S. C. Self-assembly of soft-matter quasicrystals and their approximants. *Proc. Natl Acad. Sci. USA* **108**, 20935–20940 (2011).
- Haji-Akbari, A. *et al.* Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra. *Nature* **462**, 773–777 (2009).
- Damasceno, P. F., Engel, M. & Glotzer, S. C. Predictive self-assembly of polyhedra into complex structures. *Science* **337**, 453–457 (2012).
- Steinhardt, P. J. & Taylor, P. Methods and apparatus for eliminating moiré interference using quasiperiodic patterns. US Patent Number 4,894,726 (16 January 1989).
- Zoorob, M. E., Charlton, M. D. B., Parker, G. J., Baumberg, J. J. & Netti, M. C. Complete photonic bandgaps in 12-fold symmetric quasicrystals. *Nature* **404**, 740–743 (2000).
- Grünbaum, B. & Shephard, G. C. *Tilings and Patterns* Ch. 10 (Freeman, 1986).
- Barkan, K., Diamant, H. & Lifshitz, R. Stability of quasicrystals composed of soft isotropic particles. *Phys. Rev. B* **83**, 172201 (2011).
- Skibinsky, A., Buldyrev, S. V., Scala, A., Havlin, S. & Stanley, H. E. Quasicrystals in a monodisperse system. *Phys. Rev. E* **60**, 2664–2669 (1999).
- Jagla, E. A. Phase behavior of a system of particles with core collapse. *Phys. Rev. E* **58**, 1478–1486 (1998).
- Engel, M. & Trebin, H.-R. Self-assembly of monatomic complex crystals and quasicrystals with a double-well interaction potential. *Phys. Rev. Lett.* **98**, 225505 (2007).
- Lifshitz, R. & Petrich, D. M. Theoretical model for Faraday waves with multiple-frequency forcing. *Phys. Rev. Lett.* **79**, 1261–1264 (1997).
- Henley, C. L. Cell geometry for cluster-based quasicrystal model. *Phys. Rev. B* **43**, 993–1020 (1991).
- Alder, B. J. & Wainwright, T. E. Phase transition for a hard-sphere system. *J. Chem. Phys.* **27**, 1208–1209 (1957).
- Young, D. A. & Alder, B. J. Melting curve extrema from a repulsive “step” potential. *Phys. Rev. Lett.* **38**, 1213–1216 (1977).
- Malescio, G. & Pellicane, G. Stripe phases from isotropic repulsive interactions. *Nature Mater.* **2**, 97–100 (2003).
- Phillips, C. L. & Voth, G. A. Discovering crystals using shape matching and machine learning. *Soft Matter* **9**, 8552–8568 (2013).
- Oxborrow, M. & Henley, C. L. Random square-triangle tilings: a model for twelvefold-symmetric quasicrystals. *Phys. Rev. B* **48**, 6966–6998 (1993).
- Kawamura, H. Entropy of the random triangle-square tiling. *Physica A* **177**, 73–78 (1991).
- Langon, F., Billard, L. & Chaudhari, P. Thermodynamical properties of a two-dimensional quasi-crystal from molecular dynamics calculations. *Europhys. Lett.* **2**, 625–629 (1986).
- Polik, W. F. & Burchard, W. Static light scattering from aqueous poly(ethylene oxide) solutions in the temperature range 20–90°C. *Macromolecules* **16**, 978–982 (1983).
- Motornov, M., Roiter, Y., Tokarev, I. & Minko, S. Stimuli-responsive nanoparticles, nanogels and capsules for integrated multifunctional intelligent systems. *Prog. Polym. Sci.* **35**, 174–211 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Dolinšek, M. Engel, S. Förster, N. Fujita, M. A. Glaser, C. L. Henley, G. Kahl, R. D. Kamien, Y. Kimura, R. Lifshitz, C. N. Likos, T. C. Lubensky, Y. Sakamoto, M. Schmiedeberg, A. Šiber and A. Takano for discussions. This work was supported by the Japan Society for the Promotion of Science through Grant-in-Aid for Scientific Research (C) (grant number 22540375), by the Slovenian Research Agency (grant number P1-0055) and by the Marie-Curie Initial Training Network COMPOIDS under FP7-PEOPLE-ITN-2008 (grant number 234810).

Author Contributions T.D. and P.Z. conceived the project, T.D. and T.O. performed simulations, and P.Z. proposed the tiling theory. T.D. and P.Z. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.D. (dotera@phys.kindai.ac.jp).

A two-fold increase of carbon cycle sensitivity to tropical temperature variations

Xuhui Wang¹, Shilong Piao^{1,2}, Philippe Ciais^{1,3}, Pierre Friedlingstein⁴, Ranga B. Myneni⁵, Peter Cox⁴, Martin Heimann⁶, John Miller^{7,8}, Shushi Peng¹, Tao Wang^{1,3}, Hui Yang¹ & Anping Chen⁹

Earth system models project that the tropical land carbon sink will decrease in size in response to an increase in warming and drought during this century, probably causing a positive climate feedback^{1,2}. But available data^{3–5} are too limited at present to test the predicted changes in the tropical carbon balance in response to climate change. Long-term atmospheric carbon dioxide data provide a global record that integrates the interannual variability of the global carbon balance. Multiple lines of evidence^{6–8} demonstrate that most of this variability originates in the terrestrial biosphere. In particular, the year-to-year variations in the atmospheric carbon dioxide growth rate (CGR) are thought to be the result of fluctuations in the carbon fluxes of tropical land areas^{6,9,10}. Recently, the response of CGR to tropical climate interannual variability was used to put a constraint on the sensitivity of tropical land carbon to climate change¹⁰. Here we use the long-term CGR record from Mauna Loa and the South Pole to show that the sensitivity of CGR to tropical temperature interannual variability has increased by a factor of 1.9 ± 0.3 in the past five decades. We find that this sensitivity was greater when tropical land regions experienced drier conditions. This suggests that the sensitivity of CGR to interannual temperature variations is regulated by moisture conditions, even though the direct correlation between CGR and tropical precipitation is weak⁹. We also find that present terrestrial carbon cycle models do not capture the observed enhancement in CGR sensitivity in the past five decades. More realistic model predictions of future carbon cycle and climate feedbacks require a better understanding of the processes driving the response of tropical ecosystems to drought and warming.

Climate variability related to El Niño/Southern Oscillation has a dominant role in forcing year-to-year variation in CGR, particularly the forcing of temperature variability^{9–13} (Fig. 1a and Extended Data Fig. 1). The observed positive correlation between CGR and temperature reflects the direct impacts of temperature variations in driving variations of tropical carbon fluxes^{9,10}, rather than, in reverse, the greenhouse effect of atmospheric CO₂ (Methods). Because tropical temperatures are thought to be close to the optimal photosynthetic temperature¹⁴, rising tropical temperatures could decrease vegetation photosynthesis and increase ecosystem respiration, which amplifies the influence of temperature on ecosystem carbon exchange⁹. However, vegetation productivity and respiration both increase with higher precipitation and thus offset each other⁹. These processes lead to a weaker correlation of CGR with precipitation⁹ (with coefficient $R = -0.19$; $P > 0.10$) than that of CGR with temperature, on interannual timescale. An application of causality analysis (convergent cross-mapping¹⁵) also shows that temperature is the major climatic driver of CGR variations (Extended Data Fig. 2). Below, by analysing the atmospheric CO₂ record, we investigated how the sensitivity of tropical carbon fluxes to tropical land surface mean annual temperature (MAT) has changed over the past five decades (Methods).

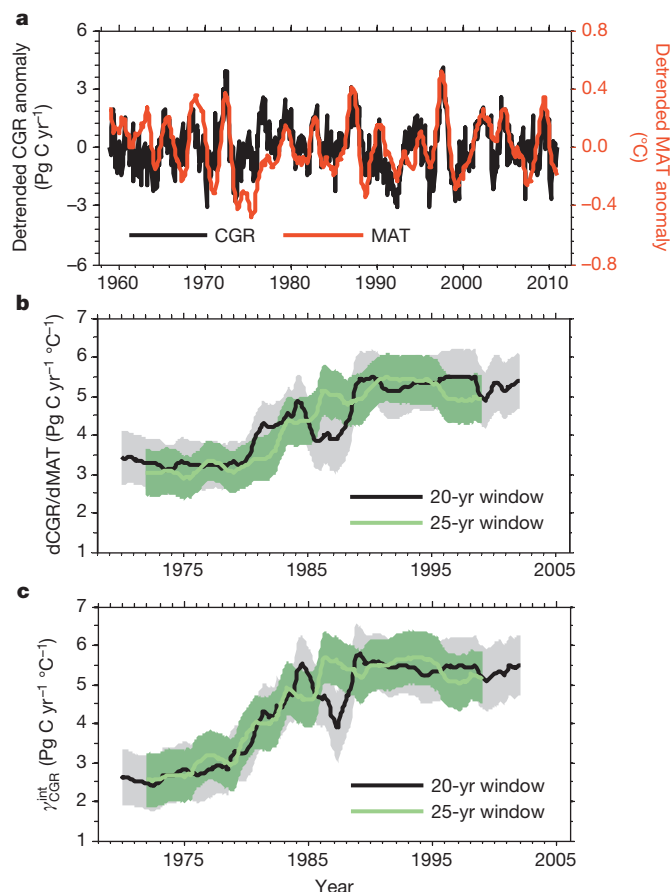


Figure 1 | Change in detrended anomalies in CGR and tropical MAT, in $dCGR/dMAT$ and in γ_{CGR}^{int} over the past five decades. **a**, Change in detrended CGR anomalies at Mauna Loa Observatory (black) and in detrended tropical MAT anomalies (red) derived from the CRU data set¹⁶. Tropical MAT is calculated as the spatial average over vegetated tropical lands ($23^\circ N$ to $23^\circ S$). The highest correlations between detrended CGR and detrended tropical MAT are obtained when no time lags are applied ($R = 0.53$, $P < 0.01$). **b**, Change in $dCGR/dMAT$ during the past five decades. **c**, Change in γ_{CGR}^{int} during the past five decades. In **b** and **c**, different colours show $dCGR/dMAT$ or γ_{CGR}^{int} estimated with moving time windows of different lengths (20 yr and 25 yr). Years on the horizontal axis indicate the central year of the moving time window used to derive $dCGR/dMAT$ or γ_{CGR}^{int} (for example, 1970 represents period 1960–1979 in the 20-yr time window). The shaded areas show the confidence interval of $dCGR/dMAT$ and γ_{CGR}^{int} , as appropriate, derived using 20-yr or 25-yr moving windows in 500 bootstrap estimates.

¹Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. ²Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100085, China. ³Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, 91191 Gif-sur-Yvette, France. ⁴College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK. ⁵Department of Earth and Environment, Boston University, Boston, Massachusetts 02215, USA. ⁶Max Planck Institute for Biogeochemistry, 07701 Jena, Germany. ⁷Global Monitoring Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, Colorado 80305, USA. ⁸Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA. ⁹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA.

The linear regression slope between CGR and tropical MAT (dCGR/dMAT) indicates the sensitivity of the tropical carbon cycle to climate interannual variability, and was used to constrain future projections of carbon cycle/climate feedbacks¹⁰. Using CGR at Mauna Loa Observatory and tropical MAT from the Climatic Research Unit (CRU) at the University of East Anglia¹⁶, we first calculated how dCGR/dMAT changes with time using a 20-yr moving window between 1960 and 2011 (both variables detrended; Fig. 1b). Within each moving window, there is a significant positive correlation between CGR and MAT ($R = 0.49\text{--}0.71$, $P < 0.01$). Furthermore, dCGR/dMAT increased from $3.4 \pm 0.4 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ (all uncertainties are s.e.; Pg C, petagrams of carbon) during 1960–1979 to $5.4 \pm 0.4 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ during 1992–2011 (Fig. 1b).

When interpreting this observed long-term increase of dCGR/dMAT, it might be argued that it does not reflect the change in the true response of CGR to temperature variations, but could be due to indirect effects of other climate variables that tend to covary with temperature, such as precipitation and solar radiation. Therefore, to isolate the role of temperature better, we defined the interannual temperature sensitivity of CGR ($\gamma_{\text{CGR}}^{\text{int}}$) as the partial derivative of CGR with respect to MAT in a multiple regression of CGR against MAT, precipitation and solar radiation (all variables detrended; Methods). The increase in dCGR/dMAT was found to reflect mainly an enhanced $\gamma_{\text{CGR}}^{\text{int}}$, rather than a change induced by covariations with the other climate variables (Extended Data Fig. 3). As shown in Fig. 1c, the value of $\gamma_{\text{CGR}}^{\text{int}}$ has increased by a factor of two, from $2.7 \pm 0.4 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ to $5.5 \pm 0.4 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ between 1960–1979 and 1992–2011. The window-by-window changes in $\gamma_{\text{CGR}}^{\text{int}}$ were fitted with three different models (the Mann–Kendall trend test, a regime shift model and a linear trend model), all of which showed that $\gamma_{\text{CGR}}^{\text{int}}$ has increased over the past five decades. However, there are too few degrees of freedom, as a result of data overlapping in neighbouring time windows. Thus, we focus below on the difference in $\gamma_{\text{CGR}}^{\text{int}}$ between the earliest time window (the first 20–25 yr) and the most recent time window (the past 20–25 yr), which are derived from fully independent subsets of the data (Methods).

We assessed the robustness of the inferred enhancement in $\gamma_{\text{CGR}}^{\text{int}}$ between the first and the last 20–25-yr period of the Mauna Loa CO₂ records. Given the significant disturbance caused by the Mount Pinatubo eruption to the global carbon cycle¹⁷, we first proposed that the cooler post-Pinatubo years characterized by enhanced carbon sinks were an unusual response of carbon sinks to MAT, which could alone explain the higher $\gamma_{\text{CGR}}^{\text{int}}$ value of the most recent two decades. Thus, we excluded the two years following the Mount Pinatubo eruption (1992–1993) from our calculation of $\gamma_{\text{CGR}}^{\text{int}}$, which had very little effect on its value, during the most recent two decades ($5.4 \pm 0.4 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$) (Extended Data Fig. 4a). To investigate whether extreme years could explain the enhancement in $\gamma_{\text{CGR}}^{\text{int}}$, we performed 500 bootstrap analyses by randomly selecting a subset of data in each 20-yr moving window to calculate $\gamma_{\text{CGR}}^{\text{int}}$. The confidence interval of these estimates is shown in Fig. 1c and confirms that the increase in $\gamma_{\text{CGR}}^{\text{int}}$ is not particularly sensitive to a few extreme years. For instance, we still obtain a significant increase in $\gamma_{\text{CGR}}^{\text{int}}$ (by a factor of 2.3–2.6) when excluding the two highest positive anomalies in CGR (Fig. 1a), which correspond to the El Niño events of 1972–1973 and 1997–1998, respectively (Extended Data Fig. 4b).

We further verified that the enhancement in $\gamma_{\text{CGR}}^{\text{int}}$ over the past five decades is robust with respect to (1) the length of the moving window for calculating $\gamma_{\text{CGR}}^{\text{int}}$ (ranging from 20 to 25 yr); (2) the application of a frequency decomposition method by singular spectrum analysis¹⁸ to separate the interannual-timescale component in CGR and MAT, instead of linear detrending (this resulted in $\gamma_{\text{CGR}}^{\text{int}}$ increasing by a factor of 1.7 for the 25-yr time windows and by a factor of 2.0 for the 20-yr time windows; Extended Data Fig. 4c); (3) the choice of climate data, using MAT from the Global Historical Climate Network¹⁹, precipitation from Global Precipitation Climatology Centre²⁰ and data on short-wave radiation²¹ ($\gamma_{\text{CGR}}^{\text{int}}$ increased by a factor of 1.9–2.1; Extended Data Fig. 4d); and (4) the choice of the other long-term CO₂ record from the South

Pole (Methods) instead of Mauna Loa ($\gamma_{\text{CGR}}^{\text{int}}$ increased by a factor of 1.7–1.9; Extended Data Fig. 4e). Overall, by analysing the histogram distribution of $\gamma_{\text{CGR}}^{\text{int}}$ values obtained with different data sources and different sensitivity tests described above (Methods), we conclude that $\gamma_{\text{CGR}}^{\text{int}}$ has robustly increased from $2.6 \pm 0.5 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ to $4.8 \pm 1.0 \text{ Pg C yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$ (Fig. 2a). This represents an average relative increase by a factor of 1.9 ± 0.3 (Fig. 2b). This enhancement is very unlikely to have resulted by chance ($< 10^{-4}$).

Finally, we applied the same analysis using the land sink calculated as a residual from all other terms of the global carbon budget, as an alternative to CGR for measuring variations in the global land carbon balance²², and also verified a doubling of $\gamma_{\text{Land Sink}}^{\text{int}}$ (Extended Data Fig. 4f). This test demonstrates that enhanced $\gamma_{\text{CGR}}^{\text{int}}$ is not explained by any known change in the variability of fossil fuel fluxes, ocean fluxes or emission from land-use change.

Tropical lands have experienced increasing drought in the past five decades^{23,24}, particularly from the mid 1970s to the early 1990s (Extended Data Fig. 5), when the increase in $\gamma_{\text{CGR}}^{\text{int}}$ began (Fig. 1c). On the regional scale, the more recent period (1992–2011) is drier than the earlier period (1960–1979), mainly in northern and central Africa and eastern South America (Extended Data Fig. 6), where a significant correlation between local MAT variability and CGR is observed (Extended Data Fig. 1). Large uncertainties exist in the spatiotemporal patterns of tropical drought changes, as indicated by the spread in the temporal evolution of three drought indices (the annual Palmer drought severity index²³ (PDSI),

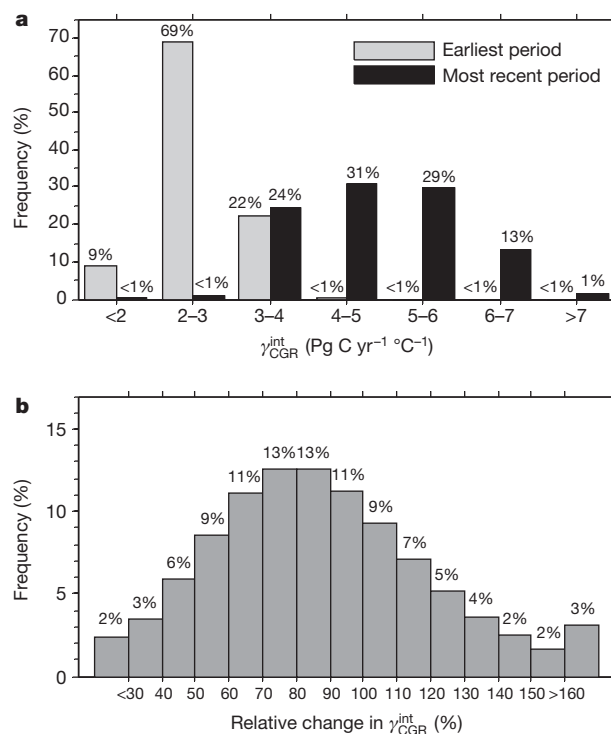


Figure 2 | Histograms of $\gamma_{\text{CGR}}^{\text{int}}$ during the earliest period and during the most recent period, and of relative change in $\gamma_{\text{CGR}}^{\text{int}}$ over the past five decades. a, Early–recent comparison; b, relative change. The relative change in $\gamma_{\text{CGR}}^{\text{int}}$ is the ratio of the calculated difference in $\gamma_{\text{CGR}}^{\text{int}}$ between the most recent period (1992–2011 for a 20-yr time-window) and the earliest period (1960–1979 for 20-yr time window) to the value of $\gamma_{\text{CGR}}^{\text{int}}$ during the earliest period. The histograms of $\gamma_{\text{CGR}}^{\text{int}}$ and of relative changes in $\gamma_{\text{CGR}}^{\text{int}}$ are calculated using bootstrap analyses, considering different data sources (CGR from Mauna Loa or the South Pole, MAT from CRU¹⁶ or the Global Historical Climate Network¹⁹, precipitation from CRU¹⁶ or the Global Precipitation Climatology Centre²⁰, and solar radiation from CRU–NCEP, ref. 21 and cloud fraction from CRU¹⁶ as a surrogate), different moving-window lengths (20–25 yr) to calculate the $\gamma_{\text{CGR}}^{\text{int}}$, and different methods to filter out the long-term increase in CGR with time (linear regression and singular spectrum analysis¹⁸).

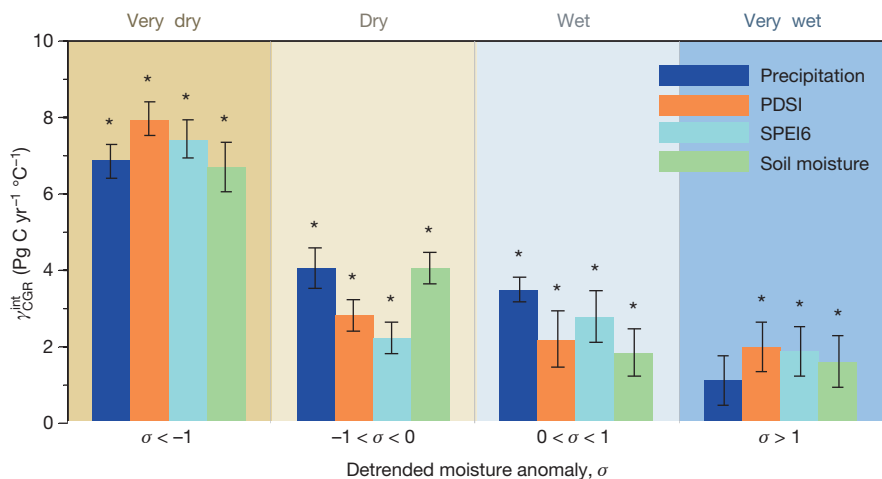


Figure 3 | $\gamma_{\text{CGR}}^{\text{int}}$ for each bin of detrended tropical moisture anomaly.

Two different drought indicators (PDSI and SPEI calculated on the 6-month scale (SPEI6)), ORCHIDEE-estimated soil moisture and precipitation were used. We grouped the 50 years' worth of data into four distinct bins of detrended tropical moisture anomaly: very dry (σ less than -1), dry (σ between

the standardized precipitation–evapotranspiration index²⁴ (SPEI) and the soil moisture estimated from the ORCHIDEE model²⁵; Extended Data Figs 5 and 6). For instance, the ORCHIDEE-estimated soil moisture shows an increase since the 1990s, whereas PDSI and SPEI are comparatively flat during the same period (Extended Data Fig. 5). To gain insight into the possible impact of tropical droughts on $\gamma_{\text{CGR}}^{\text{int}}$, we grouped $\gamma_{\text{CGR}}^{\text{int}}$ into different bins of detrended tropical moisture anomaly (Methods). Figure 3 shows that $\gamma_{\text{CGR}}^{\text{int}}$ is higher when tropical regions are drier. A Monte Carlo null experiment, assuming no change in moisture conditions, indicates that this relationship between $\gamma_{\text{CGR}}^{\text{int}}$ and drought is very unlikely to occur purely by chance ($P < 0.01$; Extended Data Fig. 7). This result suggests that soil moisture may control the response of tropical terrestrial carbon fluxes to temperature fluctuations. Many tropical ecosystems are not only subject to high temperature close to the optimal temperature of photosynthesis¹⁴, but are also subject to water limitations, at least during a dry season. Below a critical threshold for plant root water uptake, soil moisture will limit any temperature-driven increase in evapotranspiration during a warmer year. Photosynthesis will thus drop during a warmer year when soil moisture is limiting during a dry period. In turn, a reduction in photosynthesis will probably result in a larger anomaly of CO_2 input to the atmosphere²⁶, because respiration usually decreases less than photosynthesis during drought. A soil-moisture-threshold-like response of photosynthesis to temperature interannual anomalies thus can qualitatively explain why $\gamma_{\text{CGR}}^{\text{int}}$ is higher during drier periods, but additional investigation is needed to confirm that this is the main mechanism. Droughts in mid latitudes (23° – 48° north) have also increased, mostly over the past decade²⁷ (Extended Data Fig. 5), but this is inconsistent with the observed change in $\gamma_{\text{CGR}}^{\text{int}}$ (Fig. 1c). Further statistical analyses suggest that changes in mid-latitude climate must have only minor impacts on the observed $\gamma_{\text{CGR}}^{\text{int}}$ change (Extended Data Fig. 8).

Next, we explored whether terrestrial carbon cycle models can capture the observed increase in $\gamma_{\text{CGR}}^{\text{int}}$ over the past 50 years. To address this issue, we calculated the temperature sensitivity of simulated tropical net biome productivity ($\gamma_{\text{mod NBP}}^{\text{int}}$) estimated by an ensemble of five terrestrial biosphere models (Methods) previously used in ref. 22 to reconstruct historical changes in the carbon budget. We found that only one model produced an enhancement in $\gamma_{\text{mod NBP}}^{\text{int}}$ over the past five decades, and that enhancement was much smaller in magnitude than the observed change (Extended Data Fig. 9). This implies that most of the models used do not correctly capture the response of tropical carbon fluxes to climate variability⁹, probably because of biases in the response of productivity or ecosystem respiration to climate, or because

–1 and 0), wet (σ between 0 and 1) and very wet (σ greater than 1) (Methods). We then calculated $\gamma_{\text{CGR}}^{\text{int}}$ for each bin of detrended moisture anomaly, using Mauna Loa CO_2 records. The error bars indicate the s.e. of $\gamma_{\text{CGR}}^{\text{int}}$ calculated using bootstrap method (Methods). * $P < 0.05$ for $\gamma_{\text{CGR}}^{\text{int}}$ being significantly different from zero.

of inaccurate representations of soil temperature and moisture in these models. Thus, the problems present models have in reproducing the observed response of the carbon cycle to climate variability on inter-annual timescales may call into question their ability to predict the future evolution of the carbon cycle and its feedbacks to climate. Yet it should also be noted that a realistic simulation of the interannual carbon cycle dynamics provides a necessary but not sufficient test of a model's performance for the next century, because, on longer timescales, additional processes such as forest dynamics and changes in soil carbon stocks with low turnover rates become important as well²⁸.

The terrestrial carbon cycle has experienced significant changes over the past five decades^{17,29,30}. In addition to the enhanced magnitude of the land carbon sink^{17,29}, the observed significant enhancement in $\gamma_{\text{CGR}}^{\text{int}}$ found in this study provides a new perspective on a possible shift in the terrestrial carbon cycle. Increased tropical droughts are found to be related to $\gamma_{\text{CGR}}^{\text{int}}$ enhancement, highlighting the potentially important role of moisture–temperature interaction in regulating the terrestrial carbon cycle. But it should also be noted that our understanding of the mechanism of $\gamma_{\text{CGR}}^{\text{int}}$ enhancement is still limited. Further studies are needed on the response of fundamental carbon cycle processes in the tropical ecosystems and of the sensitivity of ecosystem composition and structure to environmental change. New insights could be obtained from integrated studies, combining dynamic vegetation models with data from long-term manipulative field experiments. Furthermore, new observations of atmospheric CO_2 over tropical land masses from aircraft, towers and satellites may allow for more accurate correlation with climate variables than using Mauna Loa or South Pole CO_2 data.

METHODS SUMMARY

To estimate the interannual sensitivity of CGR to tropical MAT ($\gamma_{\text{CGR}}^{\text{int}}$), we use atmospheric CGR from the two longest atmospheric records at Mauna Loa and South Pole³¹, gridded MAT from the CRU¹⁶ and from Global Historical Climate Network data¹⁹, annual precipitation from CRU and from the Global Precipitation Climatology Centre²⁰, solar radiation data from CRU–NCEP and ref. 21, and cloud fraction from CRU over vegetated areas (Methods) in the tropics. The value of $\gamma_{\text{CGR}}^{\text{int}}$ was calculated as the regression coefficient of temperature in a multilinear regression of CGR variations against variations in temperature, precipitation and solar radiation over a running time window (ranging from 20 to 25 yr). Several statistical tests (Methods) were done to assess the robustness of the increase in $\gamma_{\text{CGR}}^{\text{int}}$ over the past five decades. Two different drought indicators (PDSI²³ and SPEI²⁴) and soil moisture estimated using the ORCHIDEE model²⁵ were used to investigate the effects of drought on $\gamma_{\text{CGR}}^{\text{int}}$. We grouped the 50 years of data into four distinct bins of detrended tropical moisture anomaly (standardized departure (σ) less than -1 , between -1 and 0, between 0 and 1, and more than 1). For each bin of detrended

moisture anomaly, we calculated $\gamma_{\text{CGR}}^{\text{int}}$. Finally, the outputs of five terrestrial carbon cycle models²² were analysed to test whether these models could capture the observed increase in $\gamma_{\text{CGR}}^{\text{int}}$ during the past 50 years.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 July; accepted 27 November 2013.

Published online 26 January; corrected online 12 February 2014 (see full-text HTML version for details).

- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
- Sitch, S. *et al.* Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (DGVMs). *Glob. Change Biol.* **14**, 2015–2039 (2008).
- Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
- Meir, P. & Woodward, F. I. Amazonian rain forests and drought: response and vulnerability. *New Phytol.* **187**, 553–557 (2010).
- Davidson, E. A. *et al.* The Amazon basin in transition. *Nature* **481**, 321–328 (2012).
- Baker, D. F. *et al.* TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988–2003. *Glob. Biogeochem. Cycles* **20**, GB1002 (2006).
- Lee, K., Wanninkhof, R., Takahashi, T., Doney, S. C. & Feely, R. A. Low interannual variability in recent oceanic uptake of atmospheric carbon dioxide. *Nature* **396**, 155–159 (1998).
- Alden, C. B., Miller, J. B. & White, J. W. C. Can bottom-up ocean CO₂ fluxes be reconciled with atmospheric ¹³C observations? *Tellus B* **62**, 369–388 (2010).
- Wang, W. *et al.* Variations in atmospheric CO₂ growth rates coupled with tropical temperature. *Proc. Natl Acad. Sci. USA* **110**, 13061–13066 (2013).
- Cox, P. M. *et al.* Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**, 341–344 (2013).
- Patra, P. K., Ishizawa, M., Maksyutov, S., Nakazawa, T. & Inoue, G. Role of biomass burning and climate anomalies for land-atmosphere carbon fluxes based on inverse modeling of atmospheric CO₂. *Glob. Biogeochem. Cycles* **19**, GB3005 (2005).
- Keeling, C. D., Whorf, T. P., Wahlen, M. & van der Plicht, J. Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature* **375**, 666–670 (1995).
- Adams, J. M. & Piovesan, G. Long series relationships between global interannual CO₂ increment and climate: evidence for stability and change in role of the tropical and boreal-temperate zones. *Chemosphere* **59**, 1595–1612 (2005).
- Corlett, R. T. Impacts of warming on tropical lowland rainforests. *Trends Ecol. Evol.* **26**, 606–613 (2011).
- Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
- Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**, 693–712 (2005).
- Sarmiento, J. L. *et al.* Trends and regional distributions of land and ocean carbon sinks. *Biogeosciences* **7**, 2351–2367 (2010).
- Mahecha, M. D. *et al.* Characterizing ecosystem-atmosphere interactions from short to interannual time scales. *Biogeosciences* **4**, 743–758 (2007).
- Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21**, 2283–2296 (2008).
- Schneider, U. *et al.* GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* <http://dx.doi.org/10.1007/s00704-013-0860-x> (2013).
- Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).
- Le Quééré, C. *et al.* Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2**, 831–836 (2009).
- Dai, A. Characteristics and trends in various forms of the Palmer drought severity index during 1900–2008. *J. Geophys. Res.* **116**, D12115 (2011).
- Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* **23**, 1696–1718 (2010).
- Piao, S. *et al.* Spatiotemporal patterns of terrestrial carbon cycle during the 20th century. *Glob. Biogeochem. Cycles* **23**, GB4026 (2009).
- Schwalm, C. R. *et al.* Does terrestrial drought explain global CO₂ flux anomalies induced by El Niño? *Biogeosciences* **8**, 2493–2506 (2011).
- Dai, A. Increasing drought under global warming in observations and models. *Nature Clim. Change* **3**, 52–58 (2013).
- Randerson, J. T. Climate science: global warming and tropical carbon. *Nature* **494**, 319–320 (2013).
- Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P. & White, J. W. C. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* **488**, 70–72 (2012).
- Graven, H. D. *et al.* Enhanced seasonal exchange of CO₂ by northern ecosystems since 1960. *Science* **341**, 1085–1089 (2013).
- Keeling, C. D. *et al.* in *A History of Atmospheric CO₂ and its effects on Plants, Animals, and Ecosystems* (eds Ehleringer, J. R. *et al.*) 83–113 (Springer, 2005).

Acknowledgements We thank S. Seneviratne and G. Zhang for comments and J. Gash for English editing. This study was supported by the National Natural Science Foundation of China (41125004), the National Basic Research Program of China (grant nos 2010CB950601 and 2013CB956303) and the National Youth Top-notch Talent Support Program in China. R.B.M. was funded by the NASA Earth Science Division.

Author Contributions S. Piao, P. Ciais and X.W. designed the research; X.W. performed the analysis; S. Piao, P. Ciais and X.W. drafted the paper; and P.F., R.B.M., P. Cox, M.H., J.M., S. Peng, T.W., H.Y. and A.C. contributed to the interpretation of the results and to the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Piao (slpiao@pku.edu.cn).

METHODS

Atmospheric CO₂ concentration. The measurements of monthly mean atmospheric CO₂ concentration at the Mauna Loa Observatory were obtained from the US National Oceanic and Atmospheric Administration for the period 1959 to 2011 (<http://www.esrl.noaa.gov/gmd/ccgg/trends/>) and that from the South Pole station were obtained from the Scripps Institution of Oceanography (<http://scrippsco2.ucsd.edu/>)³¹. Note that about 15% of monthly CO₂ concentrations in the South Pole station during 1959–2011, particularly during the 1960s, were gap-filled (see <http://scrippsco2.ucsd.edu> for more processing details).

Gridded climate fields. Gridded climate fields used in this study include mean annual temperature (MAT), precipitation, solar radiation, PDSI and SPEI.

Gridded MAT, annual precipitation data and cloud fraction data are from Climatic Research Unit, University of East Anglia¹⁶ (CRU TS 3.2). CRU TS 3.2 provides monthly air temperature, precipitation and other climate variables with spatial resolution of $0.5^\circ \times 0.5^\circ$ over the global land surface during the period 1901–2011. Because solar radiation is not included in the CRU TS 3.2 data set, we used cloud fraction as a surrogate for variations in solar radiation. We further obtained solar radiation from the CRU–NCEP data product (<http://dods.extra.cea.fr/store/p529viov/cruncpe/>). CRU–NCEP is a gridded climate data product combining CRU TS 3.1 and NCEP–NCAR reanalysis. It provides 6-hourly meteorological forcing fields, including solar radiation, for carbon cycle model simulations (see, for example, ref. 2). Given that the discontinuity of the assimilated observations possibly causes the spurious trends in the reanalysis data set³², the monthly mean value of incoming solar radiation from the NCEP–NCAR reanalysis was corrected in CRU–NCEP to match the empirically derived monthly solar radiation using an approach based on latitude and sunshine hours³³, which was estimated to be proportional to observed cloud fraction in CRU TS 3.1. The preparation of the CRU–NCEP data set was performed in a carbon cycle model intercomparison project². Among all the climate stations used in generating this data set, about 500 are located between 23° north and 23° south, but their spatiotemporal distribution is uneven¹⁶. For example, ground measurements are particularly sparse over uninhabited area, including the Amazon basin and the tropical rainforest area in Southeast Asia. Thus, we have also used other climate data streams, including different interpolation approaches, climate reanalysis and a combination of ground and satellite measurements, to test the robustness of the inferred increase in $\gamma_{\text{CGR}}^{\text{int}}$ to uncertainties in the climate fields. The alternative climate data sets include air temperature fields of the Global Historical Climate Network Gridded Product (GHCN v3) of the National Climatic Data Center¹⁹, which has a spatial resolution of $5^\circ \times 5^\circ$; precipitation fields from the Global Precipitation Climatology Centre²⁰ (GPCC), which have a spatial resolution of $1^\circ \times 1^\circ$; and a solar radiation data set with a spatial resolution of $1^\circ \times 1^\circ$ (ref. 21). The short-wave radiation product from ref. 21 was obtained by correcting radiation in NCEP–NCAR reanalysis product with satellite-derived SRB v3.0 radiation data³⁴. All the data streams (ground measurement interpolation, climate reanalysis, and combination of ground and satellite measurements) have advantages and disadvantages, and thus could have differences despite the general consistency^{35,36}. For example, from the late 1990s onwards the CRU data set seems to be giving higher estimates of global precipitation than does GPCC³⁷. These differences in the climate data are one of the major sources of uncertainty in $\gamma_{\text{CGR}}^{\text{int}}$ estimates (Fig. 2).

The PDSI is a widely used index of drought stress and aridity change^{27,38}. We used global PDSI data provided by University Corporation for Atmospheric Research²⁴. The PDSI data have a spatial resolution of $2.5^\circ \times 2.5^\circ$ and a temporal resolution of one month from 1850 to 2010. This PDSI data set calculates potential evapotranspiration using the Penman–Monteith equation, which gives more realistic estimates than do Thornthwaite equations^{27,37}.

The SPEI is a multiscalar drought index similar to the standardized precipitation index but involving a climatic water balance considering both precipitation and evapotranspiration²⁴. The timescale of SPEI (or the standardized precipitation index) is defined so that an n -month scale considers the water surplus or deficit accumulated over $n - 1$ previous months²⁴. Previous studies suggest that timescales ranging from 3 to 12 months are best suited to monitoring various drought types³⁹. In this study, we adopted the same 6-month scale for SPEI (SPEI6) as in ref. 40, but also performed all calculations with SPEI on timescales of 3 months (SPEI3), 9 months (SPEI9) and 12 months (SPEI12). The global SPEI data set was provided by ref. 41 with a spatial resolution of $0.5^\circ \times 0.5^\circ$ and a temporal resolution of one month from 1901 to 2011.

Carbon fluxes and soil moisture from global ecosystem models. Annual net biome productivity from 1901 to 2008 was simulated by five dynamic global vegetation models (DGVMs) (HyLand (HYL), Lund–Potsdam–Jena (LPJ), ORCHIDEE (ORC), Sheffield–DGVM (SHE) and TRIFFID (TRI)) in the comparison study of ref. 2, and land carbon sink simulations in ref. 22 (<http://dgvm.ceh.ac.uk>). Because spatially explicit output from ORCHIDEE was not available before 1972 in this data set, results from an updated ORCHIDEE simulation are used. Starting from

pre-industrial equilibrium for all ecosystem carbon pools reached after spin-up, all models were forced with CRU–NCEP climate fields and increasing atmospheric CO₂ concentration for the period 1901–2008. It is important to note that the DGVM models produce only land–atmosphere CO₂ fluxes because of changed climate and CO₂; other drivers, such as nitrogen deposition, forest demography changes and land-use change, are not considered. Four DGVM models (HYL, LPJ, SHE and TRI) have a spatial resolution of $3.75^\circ \times 2.5^\circ$, and the updated ORCHIDEE simulation has a spatial resolution of $0.5^\circ \times 0.5^\circ$.

Analyses. The annual CGR of a specific month is calculated as the difference between the CO₂ concentration in that month and in the same month the previous year. Tropical MAT, annual precipitation, annual solar radiation, annual PDSI and annual SPEI were all calculated as the spatial average over all the vegetated land area between 23° north and 23° south. The vegetated land area is defined as all grid points for which the mean annual normalized difference vegetation index during 1982–2006 was larger than 0.1. Normalized difference vegetation index data were obtained from the Global Inventory Monitoring and Modelling Studies group⁴².

Interannual variation of CGR was found to be driven by climate variations, particularly by tropical temperature, on the basis of atmospheric CO₂ observations and terrestrial carbon cycle modelling (see, for example, refs 9, 10). This finding is consistent with field observations that tropical forest growth is more closely linked to variations in temperature than to variations in precipitation and solar radiation⁴³. To improve our understanding of the climate–CGR relationship, we also applied a convergent cross-mapping method¹⁸ to detect the linkage between CGR and tropical MAT (or other climate variables).

We calculated $\gamma_{\text{CGR}}^{\text{int}}$ over a running time window (ranging from 20 to 25 yr) during the study period. Linear trends, estimated by the least-squares method, were removed from the CGR and tropical climate (MAT, annual precipitation, mean annual solar radiation) time series within each time window. Then $\gamma_{\text{CGR}}^{\text{int}}$ was calculated as the regression slope of MAT in a multilinear regression of CGR against MAT, annual precipitation and mean annual incoming solar radiation (all variables detrended). Different time-window lengths were tested, up to a maximum length of 25 yr. This avoids data overlapping between the first and the most recent time window. Considering the fact that detrended CGR and MAT anomalies have 1–2-yr autocorrelation (Extended Data Fig. 10), a minimum time-window length was chosen as 20 yr. Note that, even though $\gamma_{\text{CGR}}^{\text{int}}$ calculated with shorter time windows, such as 15 yr, could be affected by autocorrelation, our main result—that $\gamma_{\text{CGR}}^{\text{int}}$ increased by a factor of two—is robust to changes in the length of the selected time window. We also calculated $\gamma_{\text{CGR}}^{\text{int}}$ using a linear mixed model that specifically accounts for temporal autocorrelation with a first-order autocorrelation function^{44,45}. The results show that including data autocorrelation in the regression does not change the value of $\gamma_{\text{CGR}}^{\text{int}}$ (Extended Data Fig. 10).

The window-by-window change in $\gamma_{\text{CGR}}^{\text{int}}$ was fitted using three different methods: the Mann–Kendall test⁴⁶, linear regression of $\gamma_{\text{CGR}}^{\text{int}}$ against time, and a regime shift model. We noted that the data overlapping in neighbouring time windows result in too few degrees of freedom to test the statistical significance of the trend in $\gamma_{\text{CGR}}^{\text{int}}$. Thus, statistical tests were made to compare $\gamma_{\text{CGR}}^{\text{int}}$ for the earliest time window with that for the most recent time window, which are fully independent from each other.

To test whether the shift in $\gamma_{\text{CGR}}^{\text{int}}$ during the study period was an artefact of our detrending method for CGR and tropical climate time series (MAT, annual precipitation and mean annual solar radiation), we applied frequency decomposition by singular spectrum analysis (SSA; see, for example, ref. 18) to extract the inter-annual variability of CGR and climate variables. Here the monthly CGR time series was decomposed into four frequency components: a high-frequency component⁴⁷, a seasonal component, an interannual component and a long-term trend component. Thresholds separating the frequency components were set a priori and constrained by the temporal resolution and length of CGR time series. Frequency components with the variability up to 4.5 months were regarded as the high-frequency component, whereas the thresholds separating seasonal, interannual and long-term trend components were respectively set to be 20 and 90 months, which were equally spaced on the logarithm axis of the frequency domain (see ref. 18 for more detailed descriptions of SSA). The interannual component of the climate variables was obtained in the same way. Then a linear regression between interannual components of CGR and those of the climate variables was calculated to estimate $\gamma_{\text{CGR}}^{\text{int}}$.

To estimate uncertainty in $\gamma_{\text{CGR}}^{\text{int}}$, we calculated it in 500 bootstrap analyses⁴⁸, for each combination of different data sources (CGR from Mauna Loa or the South Pole, MAT from CRU or from GHCN, precipitation from CRU or from GPCP, solar radiation from CRU–NCEP and ref. 21, and cloud fraction from CRU as a surrogate for solar radiation), different moving-window lengths (20–25 yr) to calculate $\gamma_{\text{CGR}}^{\text{int}}$, and different methods for filtering the long-term increase in CGR with time (linear regression and SSA). The uncertainty of change in $\gamma_{\text{CGR}}^{\text{int}}$ was taken from the 72,000,000 sensitivity tests.

To study the linkage between change in $\gamma_{\text{CGR}}^{\text{int}}$ and tropical drought, we grouped the 50 years' worth of data into four distinct bins of detrended tropical moisture anomaly (standardized departure (σ) less than -1 , between -1 and 0 , between 0 and 1 , and greater than 1). For each bin of detrended moisture anomaly, we calculated $\gamma_{\text{CGR}}^{\text{int}}$. The detrended soil moisture (SM) index anomaly is defined as the standardized departure from average of detrended SM:

$$\sigma(i) = \frac{\text{SM}(i) - \text{Mean}(\text{SM})}{\text{Std}(\text{SM})}$$

Here $\text{SM}(i)$ is the detrended SM for the year i , and $\text{Mean}(\text{SM})$ and $\text{Std}(\text{SM})$ are the average and standard deviation of the SM during the whole study period, respectively. An alternative of binning tropical moisture anomalies into eight bins on the basis of σ (less than -1.5 , between -1.5 and -1 , between -1 and -0.5 , between -0.5 and 0 , between 0 and 0.5 , between 0.5 and 1 , between 1 and 1.5 , and greater than 1.5) did not qualitatively change the result.

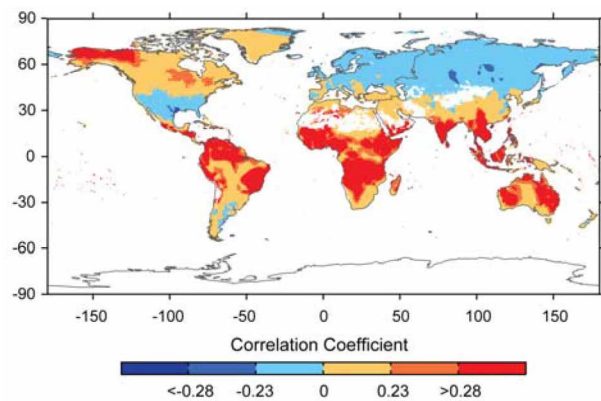
To test the robustness of the observed relationship between $\gamma_{\text{CGR}}^{\text{int}}$ and anomalies (Fig. 3), we performed a null Monte Carlo experiment. In the Monte Carlo experiment, we randomly generated artificial time series of CGR, MAT, annual precipitation and annual solar radiation following a multivariate normal distribution defined by the mean and covariations of the observed time series during 1960–2011. Because the relationship between CGR and MAT is forced to be stable during the entire period, we expected no change in $\gamma_{\text{CGR}}^{\text{int}}$ (Extended Data Fig. 7). The hypothesis to test is whether the covariations between CGR, MAT and precipitation could result in differences in $\gamma_{\text{CGR}}^{\text{int}}$ for different bins of detrended precipitation anomalies, as observed in Fig. 3.

Impacts of interannual CO_2 variations on interannual temperature variations. To estimate quantitatively the impacts of year-to-year variations in CO_2 on temperature variations, we followed refs 49, 50 to calculate the radiative forcing of CO_2 from its concentration, and used the climate sensitivity from different models. The transient temperature change due to changing CO_2 concentration is⁵⁰

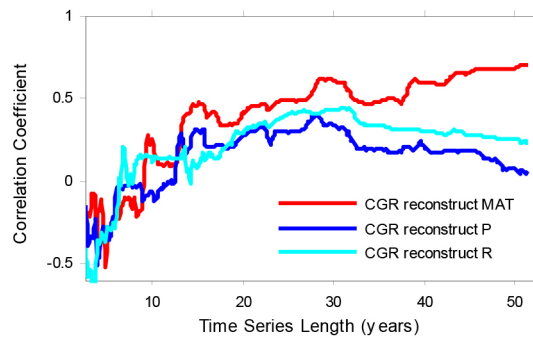
$$T_{i+1} - T_i = \frac{\alpha}{\log(2)} \log\left(\frac{C_{i+1}}{C_i}\right)$$

where T_i and C_i are the mean annual temperature and atmospheric CO_2 concentration in year i , respectively, and α is the transient temperature change in response to the doubling of CO_2 concentration. The likely range of transient climate sensitivity is $\alpha = 1.0$ – 2.5 °C (ref. 51). Here we used the largest value in this range to estimate the maximum plausible impact of CO_2 greenhouse effect on temperature variations. The result suggests that the impact of CO_2 variations on interannual temperature variations is one order of magnitude smaller than the observed interannual variations in mean annual temperature. It confirms that the impact of the CO_2 greenhouse effect on temperature is not the reason for the observed significant correlation between CGR and temperature.

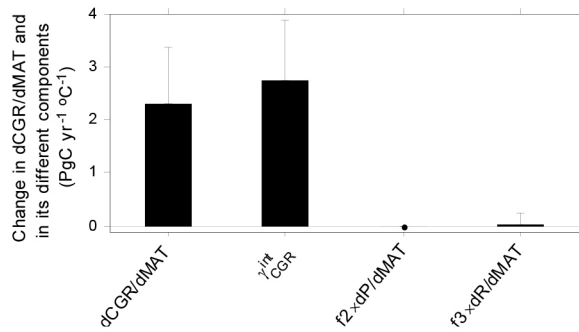
32. Kistler, R. *et al.* The NCEP–NCAR 50-year reanalysis: monthly means CD–ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**, 247–267 (2001).
33. Angstrom, A. Solar and terrestrial radiation: report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Q. J. R. Meteorol. Soc.* **50**, 121–126 (1924).
34. Stackhouse, P. *et al.* 12-year surface radiation budget data set. *GEWEX News* **14**, 10–12 (2004).
35. Dinku, T., Connor, S. J., Ceccato, P. & Ropelewski, C. F. Comparison of global gridded precipitation products over a mountainous region of Africa. *Int. J. Climatol.* **28**, 1627–1638 (2008).
36. Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* 241–265 (Cambridge Univ. Press, 2007).
37. Sheffield, J., Wood, E. F. & Roderick, M. L. Little change in global drought over the past 60 years. *Nature* **491**, 435–438 (2012).
38. Seneviratne, S. I. *et al.* Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**, 125–161 (2010).
39. Edwards, D. C. *Characteristics of 20th Century Drought in the United States at Multiple Time Scales*. MSc thesis, Colorado State Univ. (1997).
40. Hirschi, M. *et al.* Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geosci.* **4**, 17–21 (2011).
41. Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I., Angulo, M. & El Kenawy, A. A new global 0.5° gridded dataset (1901–2006) of a multiscalar drought index: comparison with current drought index datasets based on the Palmer drought severity index. *J. Hydrometeorol.* **11**, 1033–1043 (2010).
42. Tucker, C. J. *et al.* An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.* **26**, 4485–4498 (2005).
43. Clark, D. A., Piper, S. C., Keeling, C. D. & Clark, D. B. Tropical rain forest tree growth and atmospheric carbon dynamics linked to interannual temperature variation during 1984–2000. *Proc. Natl Acad. Sci. USA* **100**, 5852–5857 (2003).
44. van Mantgem, P. J. & Stephenson, N. L. Apparent climatically induced increase of tree mortality rates in a temperate forest. *Ecol. Lett.* **10**, 909–916 (2007).
45. Carnicer, J. *et al.* Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proc. Natl Acad. Sci. USA* **108**, 1474–1478 (2011).
46. Mann, H. B. Nonparametric tests against trend. *Econometrica* **13**, 245–259 (1945).
47. Mahecha, M. D., Fürst, L. M., Gobron, N. & Lange, H. Identifying multiple spatiotemporal patterns: a refined view on terrestrial photosynthetic activity. *Pattern Recognit. Lett.* **31**, 2309–2317 (2010).
48. Manly, B. *Randomization, Bootstrap and Monte Carlo Methods in Biology* 3rd edn (Chapman & Hall/CRC, 2007).
49. Solomon, S. *et al.* (eds) *Climate Change 2001: The Scientific Basis* 351–358 (Cambridge Univ. Press, 2007).
50. Hansen, J. *et al.* Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *J. Geophys. Res.* **93**, 9341–9364 (1988).
51. Alexander, L. V. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 1–33 (Cambridge Univ. Press, 2013).



Extended Data Figure 1 | Spatial distribution of the correlation coefficient between detrended CGR and MAT anomalies. CGR anomalies are from Mauna Loa Observatory and local MAT anomalies were derived from the CRU data set for the period 1960–2011. The correlation coefficients 0.23 and 0.28 are the critical thresholds at significance levels of 0.10 and 0.05 ($n = 52$), respectively.



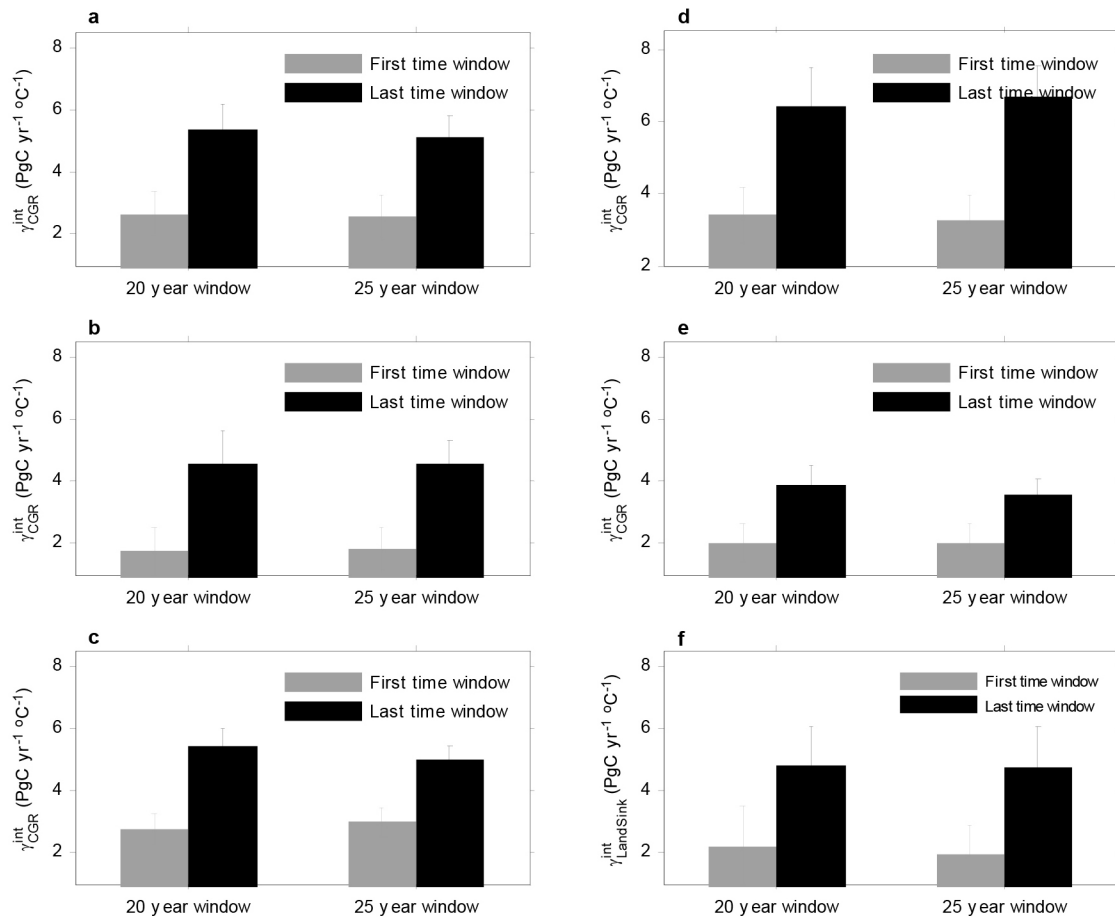
Extended Data Figure 2 | Convergent cross-mapping for reconstruction of variations in MAT, annual precipitation and mean annual solar radiation from variations in CGR. CGR data are from Mauna Loa Observatory. The CGR-reconstructed temperature curve gradually converges to a large positive correlation coefficient ($R = 0.70$), whereas the CGR-reconstructed precipitation (P) and radiation (R) curves lead to smaller correlation coefficients ($R = 0.04$ and $R = 0.23$, respectively) as time-series length increases, suggesting that CGR variations are mainly forced by temperature variations rather than by variations in precipitation and solar radiation.



Extended Data Figure 3 | Change in dCGR/dMAT, γ_{CGR}^{int} , and the effects of interannual variations of precipitation and solar radiation on the estimate of dCGR/dMAT.

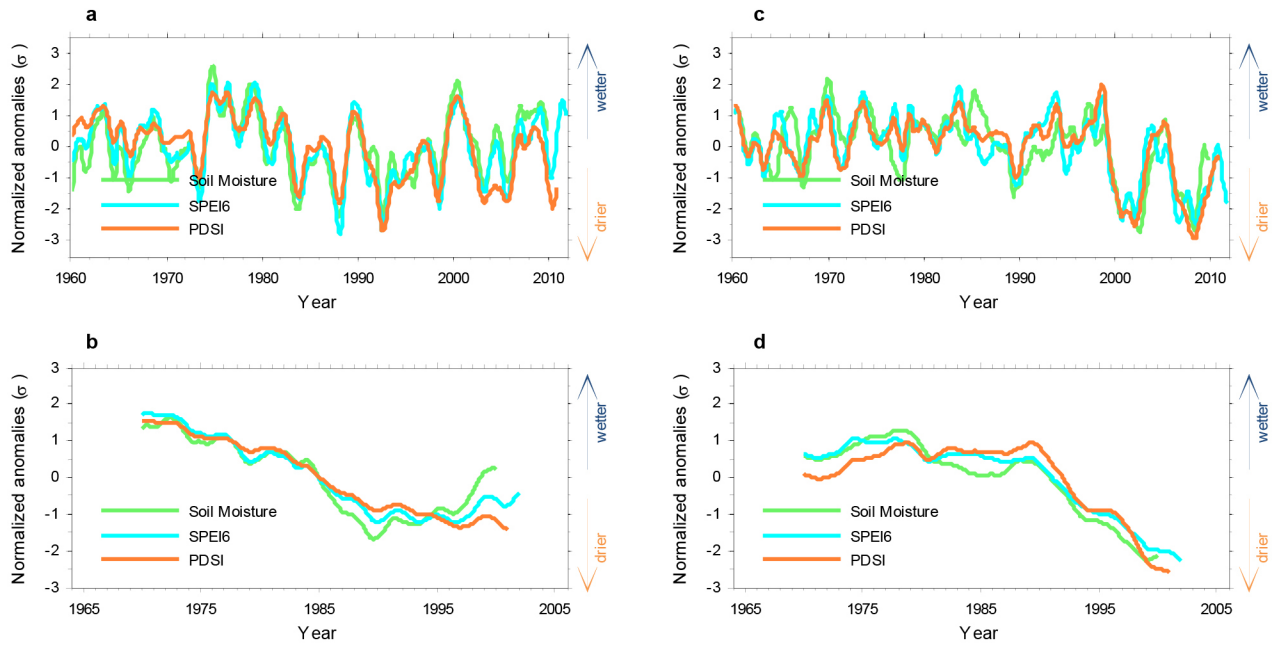
The changes are calculated between the latest two decades and the earliest two decades in 1960–2011. Precipitation and radiation effects are denoted $f2 \times dP/dMAT$ and $f3 \times dR/dMAT$, respectively.

dCGR/dMAT is calculated as the slope of MAT in the regression of CGR at Mauna Loa Observatory against MAT over the tropical vegetated land. γ_{CGR}^{int} , $f2$ and $f3$ are the slopes of MAT, precipitation and radiation, respectively, in the multiple regression of CGR against MAT, precipitation and radiation over the tropical vegetated land. $dP/dMAT$ is the slope of MAT in the regression of precipitation against MAT. $dR/dMAT$ is the slope of MAT in the regression of radiation against MAT. Error bars indicate the 95% confidence interval of the corresponding value derived from 500 bootstrap estimates.



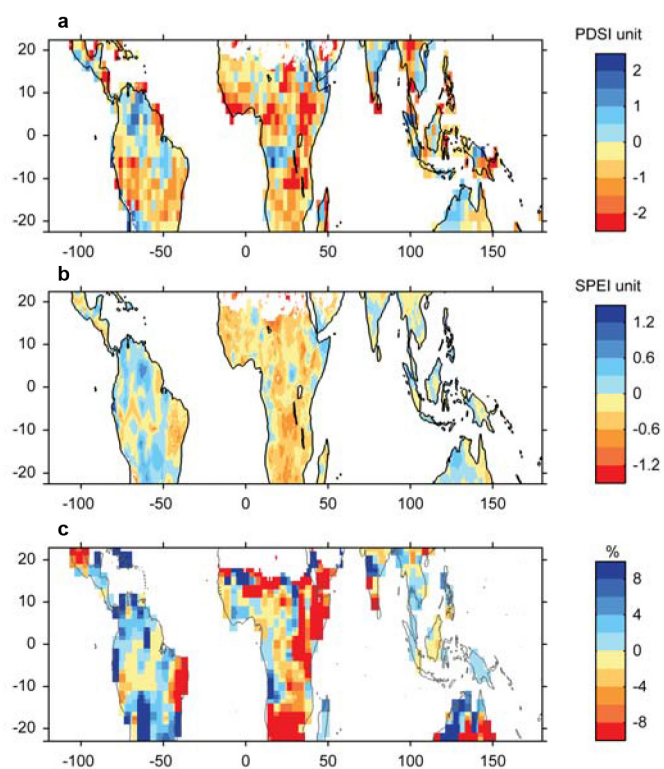
Extended Data Figure 4 | γ_{CGR}^{int} and $\gamma_{LandSink}^{int}$ for the first and the last time window during the study period 1960–2011. For 20-yr windows, the two windows are 1960–1979 and 1992–2011. For 25-yr windows, they are 1960–1984 and 1987–2011. **a**, Data from 1992 and 1993 (post-Pinatubo years) are excluded from γ_{CGR}^{int} estimates. **b**, Data from the record-high El Niño events of 1972–1973 and 1997–1998 are excluded from γ_{CGR}^{int} estimates. **c**, Interannual variations of CGR and climate variables are obtained from the frequency decomposition by SSA (Methods). **d**, γ_{CGR}^{int} is estimated with alternative climate data sets (tropical MAT is from the GHCN data set¹⁹, tropical annual

precipitation is from the GPCC²⁰ and solar radiation is from ref. 21). **e**, CGR is obtained from monthly CO₂ records at the South Pole. **f**, Interannual temperature sensitivity of the residual land carbon sink ($\gamma_{LandSink}^{int}$). The residual land carbon sink of each year is estimated from the CGR by adding the ocean sink and subtracting fossil fuel emission and emission due to land-use change²². In **a–d**, γ_{CGR}^{int} is estimated from Mauna Loa CO₂ records. Error bars indicate the 95% confidence interval of γ_{CGR}^{int} derived from 500 bootstrap estimates.

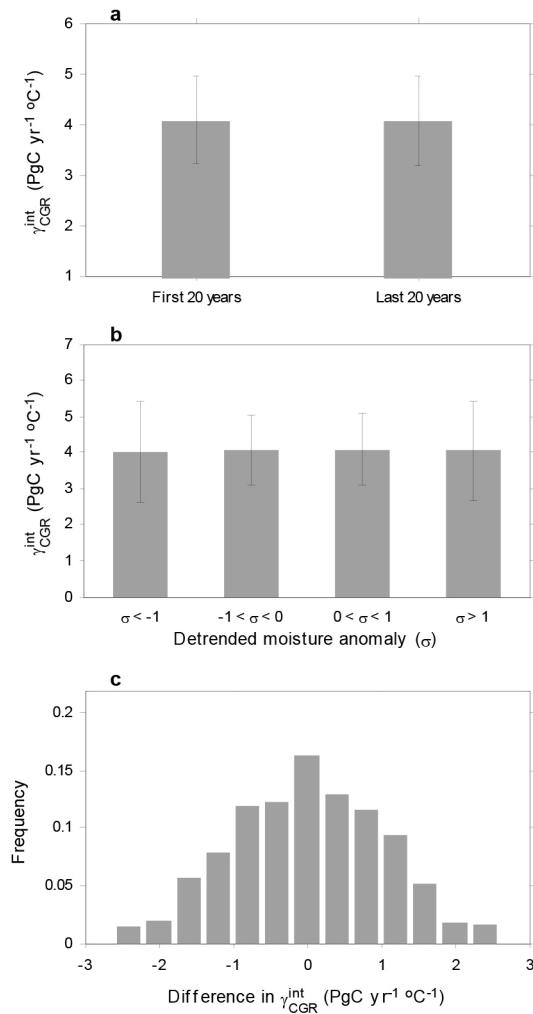


Extended Data Figure 5 | Change in drought indices and 20-yr smoothed drought indices over tropical and mid-latitude regions during the past five decades. **a**, Change in tropical (23° south to 23° north) annual PDSI, SPEI6 and ORCHIDEE-estimated soil moisture. **b**, 20-yr smoothed tropical PDSI, SPEI6 and ORCHIDEE-simulated soil moisture. **c**, Change in mid-latitude (23° north to 48° north) annual PDSI, SPEI6 and ORCHIDEE-simulated soil

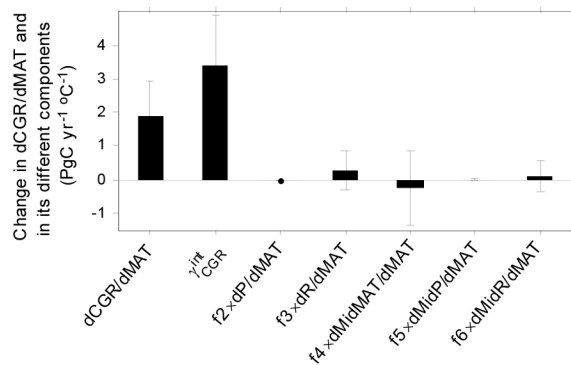
moisture. **d**, 20-yr smoothed mid-latitude PDSI, SPEI6 and ORCHIDEE-simulated soil moisture. Years on the *x*-axis indicate the central year of the 20-yr time window (for example, 1970 represents 1960–1979). All variables are normalized by their respective standard deviations. The changes in SPEI3, SPEI9 and SPEI12 are close to that in SPEI6. Note that SPEI is available till 2011, PDSI is available till 2010 and model soil moisture is available till 2009.



Extended Data Figure 6 | Spatial distribution of the difference between the latest and first 20-yr periods during the past five decades in PDSI, SPEI6 and ORCHIDEE-estimated soil moisture. a, PDSI; b, SPEI6; c, ORCHIDEE-simulated soil moisture. The changes in SPEI3, SPEI9 and SPEI12 are close to that in SPEI6.

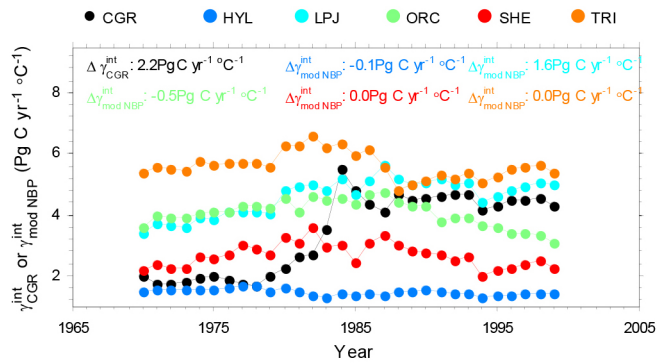


Extended Data Figure 7 | Change in γ_{CGR}^{int} and the relationship between γ_{CGR}^{int} and precipitation anomalies in the null Monte Carlo experiment. **a**, γ_{CGR}^{int} in the first and last 20 yr during 1960–2011. **b**, γ_{CGR}^{int} for each bin of detrended tropical precipitation anomalies, which are divided into four bins (standardized departure (σ) less than -1 , between -1 and 0 , between 0 and 1 , and greater than 1). Values of γ_{CGR}^{int} for different bins of detrended tropical precipitation anomaly are similar. **c**, Frequency distribution of the difference between γ_{CGR}^{int} calculated for $\sigma > 1$ (wet conditions) and γ_{CGR}^{int} calculated for $\sigma < -1$ (dry conditions). The probability of the observed difference (ranging from 5.1 to 6.0 PgC yr⁻¹ °C⁻¹; Fig. 3) occurring purely by chance is very low ($P < 0.01$). Error bars indicate the confidence intervals of the corresponding estimates.

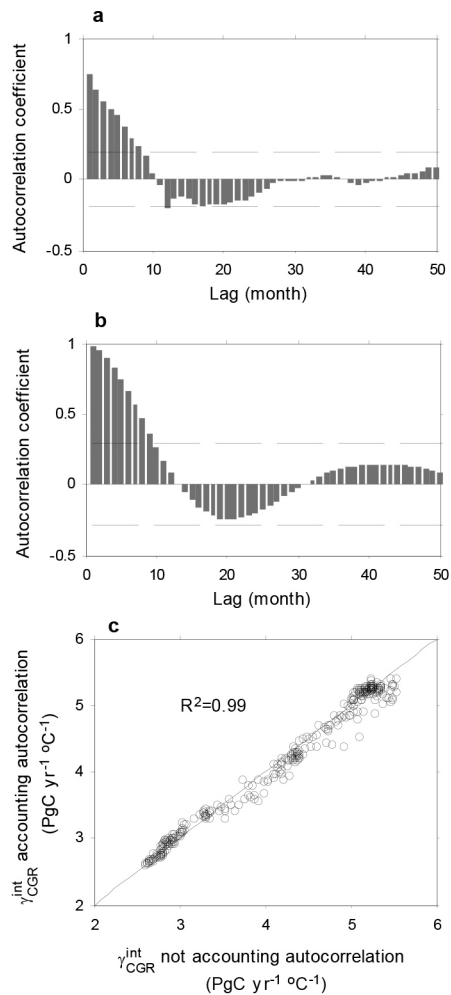


Extended Data Figure 8 | Change in dCGR/dMAT and γ_{CGR}^{int} and variational effects on estimating dCGR/dMAT between the earlier and latest two decades during 1960–2011.

The effects of interannual variations of tropical precipitation (P), tropical short-wave solar radiation (R), mid-latitude (23° north to 48° north) temperature (MidMAT), mid-latitude precipitation (MidP) and mid-latitude short-wave solar radiation (MidR) are denoted $f_2 \times dP/dMAT$, $f_3 \times dR/dMAT$, $f_4 \times dMidMAT/dMAT$, $f_5 \times dMidP/dMAT$ and $f_6 \times dMidR/dMAT$, respectively, where f_2 , f_3 , f_4 , f_5 and f_6 are respectively the slopes of P, R, MidMAT, MidP and MidR in the multiple regression of CGR against P, R, MidMAT, MidP and MidR. dy/dx represents the slope of x in the regression of y against x . Error bars indicate the 95% confidence interval of the corresponding value derived from 500 bootstrap estimates.



Extended Data Figure 9 | Change in $\gamma_{\text{CGR}}^{\text{int}}$ and $\gamma_{\text{modNBP}}^{\text{int}}$ during 1960–2008 with a 20-yr moving time window. $\gamma_{\text{modNBP}}^{\text{int}}$ is the interannual temperature sensitivity of tropical net biome productivity, estimated using five carbon cycle models (HYL, LPJ, ORC, SHE and TRI). To be consistent with model estimated annual net biome productivity, CGR of a specific year is calculated as the difference between the December Mauna Loa CO₂ concentration of the year and that of December the previous year. Positive value of $\gamma_{\text{modNBP}}^{\text{int}}$ indicates reduced anomalies of carbon sinks during warm years.



Extended Data Figure 10 | Autocorrelations in CGR and MAT, and their impacts on the estimates of $\gamma_{\text{CGR}}^{\text{int}}$. **a**, Autocorrelation coefficients for detrended anomalies of CGR from Mauna Loa during 1960–2011. **b**, Autocorrelation coefficients for detrended anomalies of MAT during 1960–2011. Dashed lines in **a** and **b** indicate 95% confidence bands. **c**, The comparison between $\gamma_{\text{CGR}}^{\text{int}}$ calculated in the multiple regression of interannual variations of the Mauna Loa CGR record against interannual variations in temperature, precipitation and solar radiation (x axis) and $\gamma_{\text{CGR}}^{\text{int}}$ calculated in the linear mixed model with same independent variables and a first-order autocorrelation function (y axis). Solid line indicates 1:1 ratio. This 1:1 relationship holds for $\gamma_{\text{CGR}}^{\text{int}}$ derived for different time-window lengths.

Global conservation outcomes depend on marine protected areas with five key features

Graham J. Edgar¹, Rick D. Stuart-Smith¹, Trevor J. Willis², Stuart Kininmonth^{1,3}, Susan C. Baker⁴, Stuart Banks⁵, Neville S. Barrett¹, Mikel A. Becerro⁶, Anthony T. F. Bernard⁷, Just Berkhout¹, Colin D. Buxton¹, Stuart J. Campbell⁸, Antonia T. Cooper¹, Marlene Davey¹, Sophie C. Edgar⁹, Günter Försterra¹⁰, David E. Galván¹¹, Alejo J. Irigoyen¹¹, David J. Kushner¹², Rodrigo Moura¹³, P. Ed Parnell¹⁴, Nick T. Shears¹⁵, German Soler¹, Elisabeth M. A. Strain¹⁶ & Russell J. Thomson¹

In line with global targets agreed under the Convention on Biological Diversity, the number of marine protected areas (MPAs) is increasing rapidly, yet socio-economic benefits generated by MPAs remain difficult to predict and under debate^{1,2}. MPAs often fail to reach their full potential as a consequence of factors such as illegal harvesting, regulations that legally allow detrimental harvesting, or emigration of animals outside boundaries because of continuous habitat or inadequate size of reserve^{3–5}. Here we show that the conservation benefits of 87 MPAs investigated worldwide increase exponentially with the accumulation of five key features: no take, well enforced, old (>10 years), large (>100 km²), and isolated by deep water or sand. Using effective MPAs with four or five key features as an unfished standard, comparisons of underwater survey data from effective MPAs with predictions based on survey data from fished coasts indicate that total fish biomass has declined about two-thirds from historical baselines as a result of fishing. Effective MPAs also had twice as many large (>250 mm total length) fish species per transect, five times more large fish biomass, and fourteen times more shark biomass than fished areas. Most (59%) of the MPAs studied had only one or two key features and were not ecologically distinguishable from fished sites. Our results show that global conservation targets based on area alone will not optimize protection of marine biodiversity. More emphasis is needed on better MPA design, durable management and compliance to ensure that MPAs achieve their desired conservation value.

A multitude of socio-economic and biological factors influence the responses of species to protection within MPA networks, adding considerable uncertainty when making specific predictions regarding the conservation benefits of new MPAs. Even within well-designed MPAs, populations of marine species can respond quite differently to prohibitions on fishing as a consequence of species-specific factors such as mobility, larval dispersal, fecundity, longevity, indirect interactions among species, environmental context, and overall level of exploitation before protection^{5,6}. To assess the extent to which MPAs fulfil their ecological potential, we used a database unprecedented in geographic scale to investigate how conservation value, characterized by ecological response of fish communities within MPAs, is affected by the cumulative effects of five key planning and management features: (1) degree of fishing permitted within MPAs; (2) level of enforcement; (3) MPA age; (4) MPA size; and (5) presence of continuous habitat allowing unconstrained movement of fish across MPA boundaries^{6–10}. Although previous studies have considered these factors independently, this is the first study, to

our knowledge, that considers them simultaneously, using data collected globally with standardized methods.

Observations from the subset of MPAs that seem to work effectively—that is, they include at least four of five ‘NEOLI’ (no take, enforced, old, large and isolated) features—are additionally used to infer ecological condition associated with unfished reefs. For this aspect, we used the global network of MPAs as a vast ecological experiment, where effective no-take areas represent human predator exclusion plots within a matrix of fished coasts¹¹.

Eight community-level metrics were assessed using data from 40 nations on shallow reef fish densities and sizes provided by researchers and trained volunteer divers participating in the Reef Life Survey (RLS) programme¹². A total of 964 sites in 87 MPAs were surveyed (Extended Data Fig. 1a), with data aggregated into 121 MPA/ecoregion groupings for analysis. MPA means were compared with statistical predictions for fished coasts using data from 1,022 non-MPA sites surveyed in 76 of the 232 Marine Ecoregions of the World¹³ (Extended Data Fig. 1b and Supplementary Data Table 1). The four community metrics investigated, each widely considered to respond to MPA declaration^{14,15}, were: (1) total biomass of all fishes; (2) total biomass of large (>250 mm length) fishes; (3) species richness of all fishes (number of species sighted per transect); and (4) species richness of large fishes. We also estimated the total biomass of three commercially important taxa (sharks, groupers and jacks), with unexploited damselfishes providing a control group for effects evident on targeted fishery groups. Effect size was calculated using the log ratio of measured values in MPAs relative to values predicted using global models for fished coasts.

Among 14 environmental and socio-economic covariates used in random forest models¹⁶ to develop predictions for fished coasts, mean sea surface temperature, annual temperature range, photosynthetically active radiation, and latitude consistently exerted the strongest influence on the global distribution of species richness and biomass metrics (Extended Data Fig. 2). Biomass of groupers and jacks was also greatly influenced by human population density, and the biomass of sharks and groupers was influenced by phosphate concentration.

Fish species richness along fished coasts peaked in the southeast Asian ‘coral triangle’ region (Fig. 1a), as expected^{12,17}. However, when only the number of large fishes sighted along transects was considered (Fig. 1b), the global centre of species richness shifted to more isolated locations within the Indo-Pacific region. Overfishing of large predatory fishes presumably contributed to these geographical patterns. Sharks, groupers and other large fishes were present within the coral triangle

¹Institute for Marine and Antarctic Studies, University of Tasmania, GPO Box 252-49, Hobart, Tasmania 7001, Australia. ²Institute of Marine Sciences, School of Biological Sciences, University of Portsmouth, Ferry Road, Portsmouth PO4 9LY, UK. ³Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, SE-106 91 Stockholm, Sweden. ⁴School of Plant Science, University of Tasmania, GPO Box 252, Hobart, Tasmania 7001, Australia. ⁵Charles Darwin Foundation, Puerto Ayora, Galapagos, Ecuador. ⁶The Bites Lab, Natural Products and Agrobiological Institute (IPNA-CSIC), 38206 La Laguna, Tenerife, Spain. ⁷Elwandle Node, South African Environmental Observation network, Private Bag 1015, Grahamstown 6140, South Africa. ⁸Wildlife Conservation Society, Indonesia Marine Program, Jalan Atletik No. 8, Bogor Jawa Barat 16151, Indonesia. ⁹Department of Water, Perth, Western Australia 6000, Australia. ¹⁰Facultad de Recursos Naturales, Escuela de Ciencias del Mar, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile. ¹¹Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, Bvd Brown 2915, 9120 Puerto Madryn, Argentina. ¹²Channel Islands National Park, United States National Park Service, 1901 Spinnaker Dr., Ventura, California 93001, USA. ¹³Instituto de Biologia, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Rio de Janeiro 21941-902, Brazil. ¹⁴Scripps Institution of Oceanography, UC San Diego, Mail Code 0227, 9500 Gilman Dr., La Jolla, California 92093-0227, USA. ¹⁵Leigh Marine Laboratory, University of Auckland, 160 Goat Island Road, Leigh 0985, New Zealand. ¹⁶Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, Università di Bologna, Via San Alberto, Ravenna 46100, Italy.

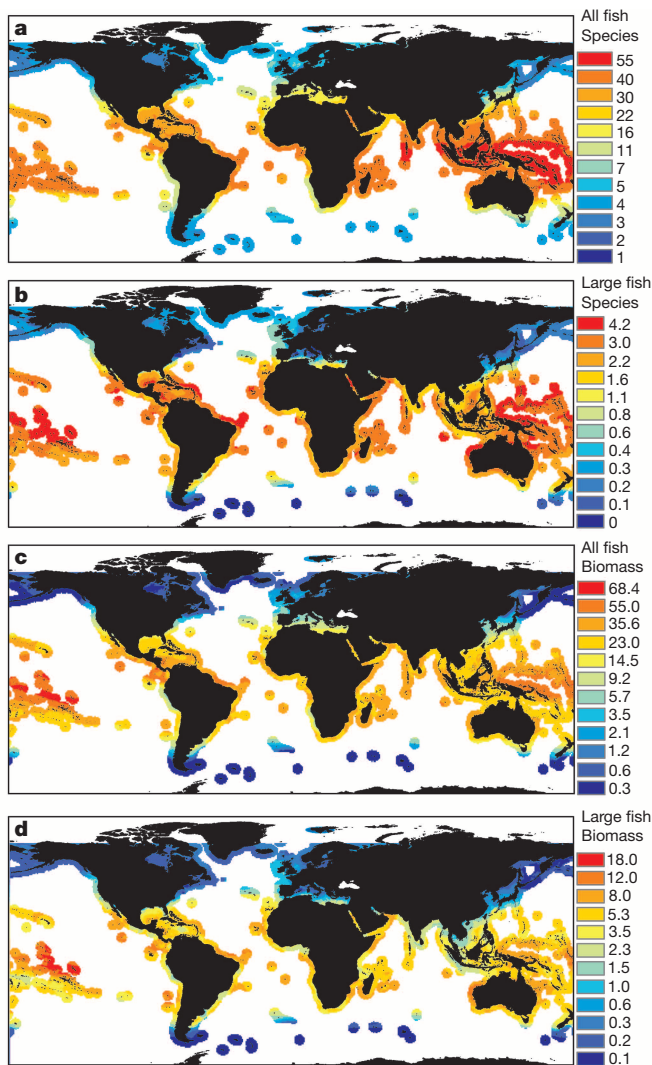


Figure 1 | Predicted global distribution of four community metrics for fishes associated with coral and rocky reefs outside of MPAs. Predictions are from random forest models developed using data from 1,022 sites in fished locations worldwide. **a**, Species richness of all fishes (number of species sighted per 250 m²). **b**, Species richness of large (>250 mm total length) fishes (per 250 m²). **c**, Total biomass of all fishes (kg per 250 m²). **d**, Total biomass of large fishes (kg per 250 m²).

region but had been exploited to near absence on most reefs, and so were rarely recorded on transects; consequently, observed species richness of large fishes was relatively low.

Our predictive models indicated that total fish and large fish biomass were highest in French Polynesia and the nearby Line Islands (Figs 1c, d), and sharks, groupers and jacks also had disproportionally high biomass in that region (Extended Data Figs 3a–c). Shark biomass on fished coasts was also very high off the Pitcairn Island group, and northeastern and northwestern Australia. Reassuringly, high shark and grouper biomass was accurately predicted for Galapagos, regardless that no data from fished sites in the oceanic tropical eastern Pacific region were used to generate the predictive models. At the time of the surveys, all islands in the region (Galapagos, Cocos and Malpelo) were within MPAs; however, data obtained before fishing restrictions in Galapagos indicate anomalously high shark and grouper biomass for fished coasts in that archipelago (S.B. and G.J.E., unpublished data). Damselfishes occurred in relatively high abundance in all tropical ocean basins (Extended Data Fig. 3d).

Across all 87 MPAs investigated, species richness of large fishes was 36% greater inside MPAs compared to fished areas (95% confidence

interval (CI), 16–60% increase), biomass of large fishes was 35% greater (CI 3–78% increase) and sharks 101% greater (CI 17–239% increase). Nevertheless, for species richness of all fishes and the other four biomass metrics investigated, no significant difference ($P > 0.05$) was found between levels observed in MPAs and those predicted for fished coasts. Moreover, many MPAs possessed fish biomass well below predicted regional averages, as indicated by the large percentage of MPAs with negative log ratios for total biomass, ranging from 25% of MPAs for large fishes to 31% for sharks to 47% for groupers. These negative values indicate considerable site-scale variability in fish densities, with some MPA sites exhibiting low fish biomass due to local habitat variability between survey sites and, in other cases, a bias resulting from stakeholder consultation processes before MPA declaration aimed at minimizing lost fishing opportunity¹⁸.

The poor overall performance of MPAs worldwide in terms of recovery of fish biomass relative to fished sites was due to a high frequency of ineffective MPAs and high spatial variability in fish densities, rather than an absence of recovery in all MPAs. The efficacy of MPAs was strongly influenced by the five NEOLI planning and management features (no take, enforced, old, large and isolated), with MPAs that scored highly with multiple NEOLI features typically having highly elevated biomass of exploitable fishes compared to fished sites (Fig. 2). MPAs with at least four NEOLI features were distributed across six countries in three oceans (Extended Data Fig. 1a) and a range of environmental conditions, indicating that model outputs and conclusions were not strongly regionally biased.

No significant differences were evident between fished sites (zero features) and MPAs with one or two NEOLI features; however, effect sizes rose rapidly when the number of features increased from three to five (Fig. 2 and Extended Data Fig. 4). For example, the measured rises in mean values within MPAs relative to fished areas for total fish biomass, total large fish biomass and shark biomass with three NEOLI features were 30%, 66% and 104%, respectively. These increases were, however, modest compared to values when all five NEOLI features were present, with large increases of 244%, 840% and 1,990%, respectively. Similar marked increases in biomass were evident for groupers (882%) and jacks (864%). Non-fished damselfishes showed a smaller mean increase of 111% at MPAs with five NEOLI features. This increase was on the margins of statistical significance, lying outside the 95% confidence interval (Extended Data Fig. 4) but nonsignificant ($P < 0.05$) when assessed with a *t*-test, which adjusts for small sample size.

All four MPAs with five NEOLI features were small oceanic islands (Cocos, Costa Rica; Malpelo, Colombia; Kermadec Islands, New Zealand; and Middleton Reef, Australia), raising a potential concern that calculated effect sizes were biased by plankton and pelagic fish subsidies that enlarge food webs at isolated oceanic locations. ‘Oceanic island’ was, however, included as a categorical covariate in random forest models, therefore model predictions should accommodate small island effects. Regardless, further investigation into the contribution of external subsidies to food webs at isolated MPAs is warranted. Alternative explanations for elevated damselfish numbers in the most effective MPAs compared with poorly protected MPAs include reduced fishing-related habitat deterioration such as dynamite damage to coral, and trophic cascades involving smaller predators that consume damselfishes and are prey to sharks and groupers.

No-take regulations, efficient enforcement, large area (>100 km²) and old age (>10 years) each contributed similar increases in fish biomass within MPAs (Fig. 2). However, isolation, a categorical factor that distinguished MPAs with reef habitat surrounded by deep (>25 m) water or large expanses of sand from MPAs with shallow reef habitat extending to fished areas, seemed to exert a stronger influence for community-level biomass and richness metrics than the other four features. For example, the mean increase (95% CI) for total fish biomass associated with MPAs with three NEOLI features was 100% (14–252%) when one of the three features was isolation, compared to 14% (–18%–58%) for three NEOLI MPAs when isolation was not included. Compliance

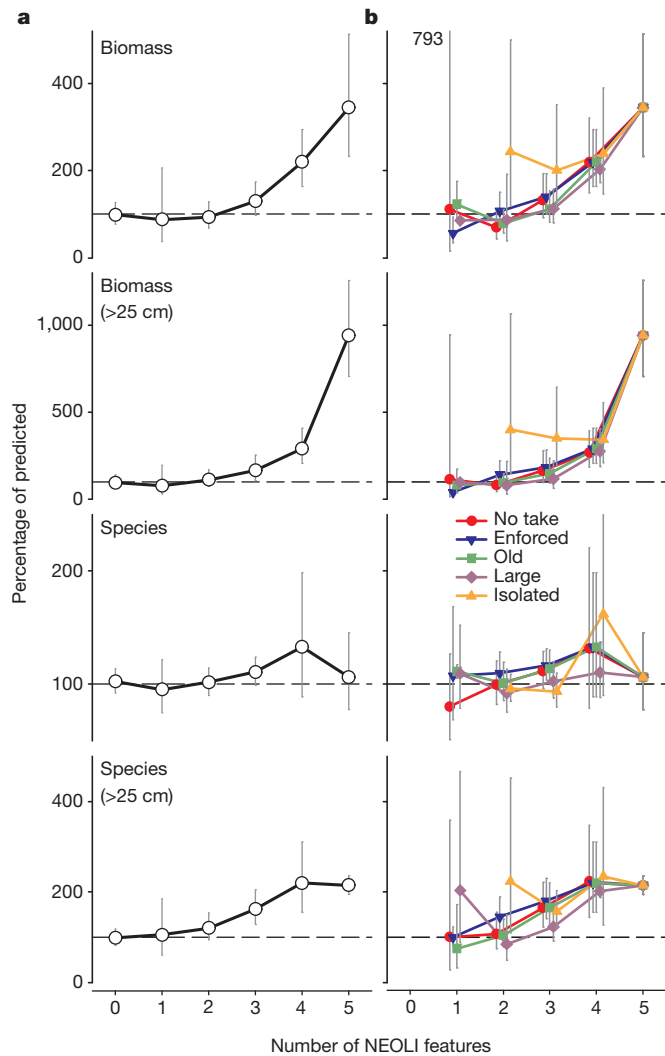


Figure 2 | Mean response ratios for MPAs with different numbers of NEOLI (no take, enforced, old, large, isolated) features. Mean ratio values have been back transformed from logs and expressed as percentages with 95% confidence intervals, with 100% equivalent to fished coasts. Sites on fished coasts have 0 NEOLI features. **a**, Mean response ratios for four community metrics. **b**, Mean response ratios for community metrics where each NEOLI feature was included within the set examined. The 'no-take' plot with two features, for example, depicts the mean response for no-take MPAs with a single other NEOLI feature. 95% confidence limits that lie off-scale are shown by number. Samples sizes are shown in Extended Data Table 1.

may have contributed to the isolation effect, in that isolated MPAs are generally well demarcated for control purposes. They are readily recognized by fishers and more easily policed than coastlines with complicated mosaics of no take, restricted take and fishing zones. Although very important, the effect of isolation was similar in magnitude—rather than clearly superior—to other MPA features for biomass of sharks, groupers and jacks (Extended Data Fig. 4).

When MPAs that are no take and well enforced are considered, differences were evident in how the other MPA features affect different components of the fish community (Fig. 3 and Extended Data Fig. 5). Total fish biomass increased significantly from low to high levels for all five MPA features, and these same trends were magnified for large fishes (Fig. 3). Regardless of general concerns that large pelagic species move such great distances that few individuals are fully protected within MPAs¹⁹, sharks and jacks seem to receive considerable protection from fishing mortality within the large, well-enforced, no-take MPAs studied here. The biomass of sharks and groupers rose exponentially when MPAs were fully isolated, and also greatly increased with area and age. The

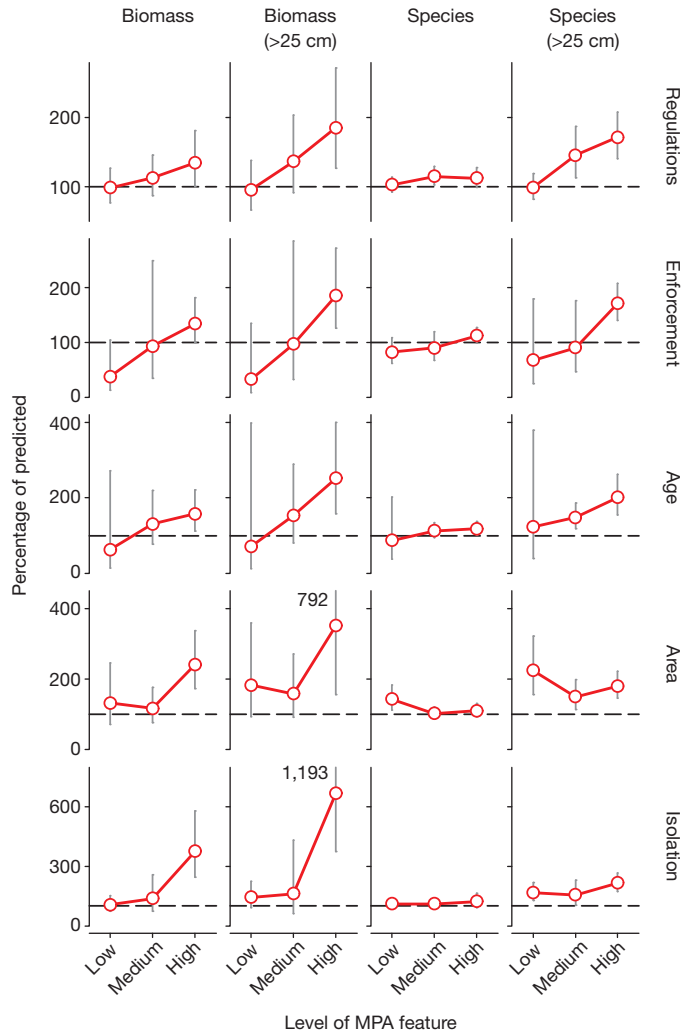


Figure 3 | Mean response ratios with 95% confidence intervals for four community metrics and low, medium and high levels of five MPA features. Values have been back transformed to per cent scale, with 100% equivalent to fished coasts. The feature 'regulations' was analysed using data from 82 MPAs with high enforcement; the feature 'enforcement' was analysed using data from 75 MPAs that are no-take; and the features 'isolation', 'age' and 'area' were analysed using data from 52 MPAs that are both no take and well enforced. 95% confidence limits that lie off-scale are shown by number. Samples sizes are shown in Extended Data Table 1.

biomass of jacks showed little isolation and age effects, but rose greatly in MPAs that were large, well enforced and no take. Damselfish biomass did not increase significantly with the accumulation of individual NEOLI features.

The large number of MPAs investigated here has allowed relatively subtle and higher order interactive MPA effects to be detected. Previous studies of MPAs have shown, for example, negligible or weak patterns associated with MPA size^{6,9,14,15,20}, and those detected here were only evident for MPAs with at least three of the NEOLI features. However, MPA size was very important for such metrics as jack biomass, which showed a stronger response to MPA area than to other metrics (Extended Data Fig. 5). This response probably resulted from time spent by actively-swimming fishes outside park boundaries, which increases probability of capture for fishes associated with small MPAs.

Species richness of large fishes exhibited a highly significant difference between MPAs with five NEOLI features and fished locations (115% increase relative to predicted, CI 95–137%; *t*-test, $P < 0.0001$; Fig. 2). By contrast, MPAs with five NEOLI features did not differ significantly in total species richness (6% increase relative to predicted)

from fished locations (t -test, $P = 0.42$; Fig. 2), nor did any of the five features individually have a clear effect on species richness (Fig. 3). Thus, total species richness along transects did not detectably increase in effective MPAs, despite the presence of additional large fish species, perhaps because of food web changes in the form of reduced presence of small fish species that comprised prey of the larger predatory species^{5,21,22}. Regardless of these transect-scale effects, species richness at regional scales probably increased in areas with a mosaic of fished and effective MPAs because of the additional presence of large fishery-targeted species within the seascape¹⁸.

Of the 87 MPAs investigated, only four possessed all five NEOLI features, whereas five MPAs possessed four features, and 39, 57 and 16 MPAs possessed three, two and one feature, respectively. The low proportion of MPAs possessing four or five NEOLI features (10%), and thus regarded here as effective, probably overstates the true proportion of effective MPAs worldwide. Our survey strategy deliberately targeted well-known and well-regarded MPAs, with most large and long-established MPAs included in this study.

Although only a small subset of MPAs are ever likely to qualify as large, most MPAs could achieve the remaining four NEOLI features. MPAs require additional time to age, and sufficient will among stakeholders, managers and politicians for increased implementation of no-fishing zones, increased levels of compliance, and extension of boundaries past the limits of reef systems or to deep water. If these could be achieved in tandem with current trends for declaration of large remote 'wilderness' MPAs^{23,24}, then conservation benefits from the global MPA network should increase markedly. However, the current base is very low with only 0.08% of the world's oceans within no-take MPAs in 2008 (ref. 25), and with opportunities for an expanded network diminishing as establishment and opportunity costs for large isolated MPAs escalate in line with human population growth^{24,26}.

By using effective MPAs as an unfished standard, our study allows the first global assessment of the magnitude of fishing effects on temperate as well as tropical reef communities. Fish biomass was greatly reduced overall, with 63% of all fish biomass, 80% of large fish biomass, 93% of sharks, 84% of groupers and 85% of jacks apparently removed from reefs by fishing.

In spite of their huge magnitude, these estimates are probably conservative because they are based on the assumption that MPAs with four or five NEOLI features provide an accurate non-fished baseline for inferring historical patterns. Yet fish populations are unlikely to have fully recovered from previous impacts of fishing in four NEOLI MPAs, which were found to be less effective than five NEOLI MPAs for some metrics. Moreover, high fishing mortality rates for sharks and wide-ranging predatory fishes outside MPAs will negatively influence total numbers within boundaries through reduced immigration rates, and further recovery of fish biomass within MPAs probably continues over much longer time spans than the 10-year threshold used here to define old MPAs¹⁸. Our estimates for effective MPAs include uncertainty associated with the low number of effective MPAs surveyed, most notably for sharks, as only five of the nine category 4 and 5 NEOLI MPAs had sharks present. Also, biomass may be overestimated because of diminished flee responses from divers of large fishes in well-enforced no-take MPAs²⁷. Regardless, fishing clearly exerts a very large and ubiquitous impact on shallow reefs.

The 80% reduction in biomass of large fishes outside effective MPAs coincides with the threshold value used by the International Union for Conservation of Nature (IUCN) to categorize species as Critically Endangered for Red List assessments²⁸. Although recognizing that application of current Red List thresholds to exploited fish stocks remains contentious²⁸, the high number of large-bodied species that together average 80% decline indicates that innumerable threatened fish species probably exist, and that effective MPAs probably have a large role in safeguarding populations of many of these species⁴. Even nations with relatively well-managed fisheries have few sharks and other large predatory coastal fishes outside well-designed and mature MPAs. Given

the huge scale of fishing impacts, the rate of fish extinctions is likely to increase greatly through this century unless a refugial network of effective MPAs exists to allow persistence of large-bodied species and associated predator-dominated food webs, and broad-scale fisheries management practices significantly improve²⁹.

METHODS SUMMARY

Surveys were based on Reef Life Survey methodology^{12,30}, with support from volunteer SCUBA divers trained individually to scientific data collection standards. All fishes observed within 50 m × 5 m transect blocks were counted, and total fish lengths estimated, during swims on adjoining blocks up one side and down the other side of 50-m lines. Each transect was set along a depth contour, with two depth contours (mean 2.4) generally surveyed at each site. Sites located within 87 MPAs were investigated, with approximately half located in Australia (36) and New Zealand (8). In total, 48 MPAs were complete no take, 18 MPAs allowed limited fishing, whereas 21 MPAs were multi-zoned with interspersed no-take and limited fishing zones. Data were compiled from 171,331 underwater abundance counts of 2,544 species in 9,544 transect blocks at 1,986 sites.

We assessed effects of five MPA features (fishing regulations, enforcement, age, area and isolation), each categorized at low, medium and high levels, on eight fish community metrics (species richness of all fishes and large (>250 mm) fishes; total biomass of all fishes, large fishes, sharks, groupers, jacks and damselfishes). The magnitudes of effects were quantified using the log ratio of observed value within the MPA to predicted value at that location if the MPA did not exist (for example, $\log[B_m/B_p]$, where B_m is measured fish biomass and B_p is biomass predicted if the site was fished). Predictions were produced using random forest procedures¹⁶, where each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. Relationships were initially established between 14 covariates (environmental and socio-economic) and measured values of the eight response metrics at fished sites. These relationships were then used, with known covariate values at each MPA, to predict each of the eight community metrics at that MPA location.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 September 2013; accepted 13 January 2014.

Published online 5 February; corrected online 12 February 2014 (see full-text HTML version for details).

- Hilborn, R., Micheli, F. & De Leo, G. A. Integrating marine protected areas with catch regulation. *Can. J. Fish. Aquat. Sci.* **63**, 642–649 (2006).
- Gell, F. R. & Roberts, C. M. Benefits beyond boundaries: the fishery effects of marine reserves. *Trends Ecol. Evol.* **18**, 448–455 (2003).
- Edgar, G. J. Does the global network of marine protected areas provide an adequate safety net for marine biodiversity? *Aquat. Conserv.* **21**, 313–316 (2011).
- Mora, C. *et al.* Coral reefs and the global network of Marine Protected Areas. *Science* **312**, 1750–1751 (2006).
- Babcock, R. C. *et al.* Decadal trends in marine reserves reveal differential rates of change in direct and indirect effects. *Proc. Natl Acad. Sci. USA* **107**, 18256–18261 (2010).
- Claudet, J. *et al.* Marine reserves: size and age do matter. *Ecol. Lett.* **11**, 481–489 (2008).
- Guidetti, P. *et al.* Italian marine reserve effectiveness: does enforcement matter? *Biol. Conserv.* **141**, 699–709 (2008).
- Le Quesne, W. J. F. Are flawed MPAs any good or just a new way of making old mistakes? *ICES J. Mar. Sci.* **66**, 132–136 (2009).
- Vandeperre, F. *et al.* Effects of no-take area size and age of marine protected areas on fisheries yields: a meta-analytical approach. *Fish. Fish.* **12**, 412–426 (2011).
- Forcada, A., Bayle-Sempere, J. T., Valle, C. & Sanchez-Jerez, P. Habitat continuity effects on gradients of fish biomass across marine protected area boundaries. *Mar. Environ. Res.* **66**, 536–547 (2008).
- Walters, C. J. & Holling, C. S. Large-scale management experiments and learning by doing. *Ecology* **71**, 2060–2068 (1990).
- Stuart-Smith, R. D. *et al.* Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539–542 (2013).
- Spalding, M. D. *et al.* Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
- Côté, I. M., Mosquera, I. & Reynolds, J. D. Effects of marine reserve characteristics on the protection of fish populations: a meta-analysis. *J. Fish Biol.* **59** (Suppl. A), 178–189 (2001).
- Halpern, B. S. The impact of marine reserves: do reserves work and does reserve size matter? *Ecol. Appl.* **13** (Suppl.), 117–137 (2003).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 15–32 (2001).
- Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
- Edgar, G. J., Barrett, N. S. & Stuart-Smith, R. D. Exploited reefs protected from fishing transform over decades into conservation features otherwise absent from seascapes. *Ecol. Appl.* **19**, 1967–1974 (2009).

19. Grüss, A., Kaplan, D. M., Guénette, S., Roberts, C. M. & Botsford, L. W. Consequences of adult and juvenile movement for marine protected areas. *Biol. Conserv.* **144**, 692–702 (2011).
20. Guidetti, P. & Sala, E. Community-wide effects of marine reserves in the Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **335**, 43–56 (2007).
21. Graham, N. A. J., Evans, R. D. & Russ, G. R. The effects of marine reserve protection on the trophic relationships of reef fishes on the Great Barrier Reef. *Environ. Conserv.* **30**, 200–208 (2003).
22. Willis, T. J. & Anderson, M. J. Structure of cryptic reef fish assemblages: relationships with habitat characteristics and predator density. *Mar. Ecol. Prog. Ser.* **257**, 209–221 (2003).
23. Pala, C. Giant marine reserves pose vast challenges. *Science* **339**, 640–641 (2013).
24. Graham, N. A. J. & McClanahan, T. R. The last call for marine wilderness? *Bioscience* **63**, 397–402 (2013).
25. Wood, L. J., Fish, L., Laughren, J. & Pauly, D. Assessing progress towards global marine protection targets: shortfalls in information and action. *Oryx* **42**, 1–2 (2008).
26. Mora, C. *et al.* Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *PLoS Biol.* **9**, e1000606 (2011).
27. Cole, R. G. Abundance, size structure, and diver-oriented behaviour of three large benthic carnivorous fishes in a marine reserve in northeastern New Zealand. *Biol. Conserv.* **70**, 93–99 (1994).
28. IUCN Standards and Petitions Working Group. *Guidelines for Using the IUCN Red List Categories and Criteria* Version 7.0 (downloaded 18 November 2013 from <http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf>, 2008).
29. Mora, C. & Sale, P. F. Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Mar. Ecol. Prog. Ser.* **434**, 251–266 (2011).
30. Edgar, G. J. & Stuart-Smith, R. D. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Mar. Ecol. Prog. Ser.* **388**, 51–62 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the many Reef Life Survey (RLS) divers who contributed to data collection. Development of the RLS data set was supported by the former Commonwealth Environment Research Facilities Program, whereas analyses were supported by the Australian Research Council, a Fulbright Visiting Scholarship (to G.J.E.), the Institute for Marine and Antarctic Studies, and the Marine Biodiversity Hub, a collaborative partnership funded under the Australian Government's National Environmental Research Program. Surveys were assisted by grants from the National Geographic Society, Conservation International, Wildlife Conservation Society, Winifred Violet Scott Trust, Tasmanian Parks and Wildlife Service, the Winston Churchill Memorial Trust, University of Tasmania, and ASSEMBLE Marine. We are grateful to the many park officers who assisted the study by providing permits and assisting with field activities, and to numerous marine institutions worldwide for hosting survey trips.

Author Contributions G.J.E. and R.D.S.-S. conceived the project; G.J.E., R.D.S.-S., M.A.B., A.T.F.B., S.C.B., S.B., S.J.C., A.T.C., M.D., S.C.E., G.F., D.E.G., A.J.I., S.K., D.J.K., R.M., G.S., E.M.A.S. and many others collected the data; G.J.E., R.J.T., T.J.W., S.K. and S.C.E. prepared figures; G.J.E. drafted the initial manuscript; all authors contributed to analyses and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.E. (g.edgar@utas.edu.au).

METHODS

Survey methodology. Standardized visual census counts were made at 1,986 sites using Reef Life Survey (RLS) methodology (see RLS methods manual 'Standardised survey procedures for monitoring rocky and coral reef ecological communities' at http://reeflifesurvey.com/files/2008/09/NEW-Methods-Manual_15042013.pdf). Divers made counts and estimates of total lengths of all fish species observed during swims at $\sim 2 \text{ m s}^{-1}$ along the centre of a 5-m-wide swathe up one side and then down the other side of 50-m transect lines. Fishes sighted in transect blocks were recorded on an underwater slate, with abundance estimates made by counting individuals of less abundant species and, in locations with high fish densities, estimating the number of more abundant species. The abundance of schooling fishes was recorded by counting a subset within the school which was combined with an estimate of the proportion of the total school. Nearly all fishes were recorded to species level, with exceptions classified at the highest taxonomic resolution possible. The use of digital photography typically allowed later identification of most unidentified species, with assistance of taxonomic experts as required.

Experienced scientists and skilled recreational divers both contributed data to the RLS programme, all divers having either substantial previous experience in fish surveys or extensive one-on-one training by R.D.S.-S. or G.J.E. To provide a major element of consistency in diver contributions at the global scale, G.J.E. and R.D.S.-S. participated in most surveys, providing 31% of all data analysed. Validation tests indicated no difference in quality or composition of data provided by volunteers participating in this programme when compared to professional biologists³⁰.

Each transect was set along a depth contour, with two depth contours generally surveyed at each site (mean of 2.4 depths per site; minimum, maximum, mean \pm s.d. depth contours surveyed: 0.1 m, 42 m, 7.5 ± 4.1 m, respectively). Sites located within 87 MPAs were investigated, with approximately half located in Australia (36) and New Zealand (8). In total, 48 MPAs were no take where all fishing was prohibited, 18 MPAs allowed limited fishing, whereas 21 MPAs were multi-zoned with interspersed no-take and limited fishing zones. Data were compiled from 171,331 underwater abundance counts of 2,544 species in 9,544 transect blocks (50 m \times 5 m).

MPA features and community metrics investigated. We assessed the influence of five MPA features on eight fish community metrics calculated using field survey data. The MPA features investigated were each categorized at three levels: low (L), medium (M) and high (H). (1) Regulations. Extent that regulations restrict fishing at survey site. L, site can be openly fished with no fishing restrictions additional to those generally applied within the state; M, site located within an MPA but with some fishing methods allowed; H, no-take area within an MPA. (2) Enforcement. Extent of compliance to regulations that restrict fishing, both through overt policing and through community support for regulations. Level was decided at the time of surveys after discussion with local park authorities, and on the basis of observations of the extent of infractions while conducting fieldwork. L, little attempt at control, a 'paper park'; M, a moderate level of policing attempted, although infractions were apparent; H, appears to be well enforced, although clandestine poaching may occur. (3) Age. Period between when regulations restricting fishing were first enacted and field surveys undertaken. L, MPA zone <5 years old; M, MPA zone 5–10 years old; H, MPA zone >10 years old. (4) Area. MPA zone area, as described in management plan or documents provided locally to users. L, <1 km²; M, 1–100 km²; H, >100 km². (5) Isolation. Degree that reef habitat surveyed is isolated by habitat boundaries from adjacent fished reef. L, shallow (<25 m) reef habitat extends continuously across MPA boundary; M, a small (1–20%) percentage of zone boundary breached by continuous shallow reef habitat; H, MPA zone isolated from fishing areas by depth (>25 m) or sand barriers of at least 20 m width.

We investigated eight community metrics. (1) Species richness of all fishes. Total number of all fish species sighted within 50 m \times 5 m transect blocks. (2) Species richness of large fishes. Total number of fish species sighted within 50 m \times 5 m transect blocks for the set of individuals observed on transects exceeding the 250 mm size class bin (that is, 300 mm size and above). (3) Total fish biomass. Total biomass of all fishes sighted in 50 \times 5 m transect blocks. Estimated by combining abundance counts with size estimates using length-weight relationships provided for total length of each fish species (in some cases genus and family) in Fishbase (<http://www.fishbase.org>). Bias in divers' perceptions of fish size underwater was additionally corrected using relationships presented in ref. 31. (4) Total biomass of large fishes. Total biomass of individuals sighted in 50 \times 5 m transect blocks that exceeded the 250 mm size class bin. (5) Total biomass of sharks. Sum of biomass of all fishes in transect that belong to orders Carcharhiniformes, Heterodontiformes, Lamniformes and Orectolobiformes. (6) Total biomass of groupers. Sum of biomass of all fishes in transect that belong to family Serranidae, genera *Dermatolepis*, *Epinephelus*, *Gracila*, *Mycteroperca*, *Paralabrax*, *Plectropomus*, *Trachypoma* and *Variola*. Small serranids such as *Pseudanthias* spp. were not considered. (7) Total biomass of jacks. Sum of biomass of all fishes in transect that belong to family Carangidae. (8) Total biomass of damselfishes. Sum of biomass of all fishes in transect that belong to family Pomacentridae.

Data aggregation. To reduce spatial confounding resulting from highly clumped distribution of sites surveyed, data were aggregated before analyses as means for each ecoregion, MPA and zone type. Thus, fished sites were aggregated as mean values for each of 76 Marine Ecoregions of the World¹³, whereas MPA data were aggregated into 121 MPA zones by ecoregion combinations. Multi-zoned MPAs contributed two data points to analyses (no-take sites and restricted fishing sites), whereas very large MPAs that extended across ecoregional boundaries (for example, Great Barrier Reef Marine Park, Galapagos Marine Reserve) were also partitioned with aggregated data from each ecoregion.

Global models. Models were developed using random forest procedures¹⁶, as available in the 'extendedForest' packages for R (<https://r-forge.r-project.org/projects/gradientforest/>), to predict the distribution of the eight community metrics in inshore habitats globally, including the MPA locations investigated. Each random forest consisted of numerous (2,000 in this case) regression trees, where each tree is fit to a bootstrap sample of the biological data using a recursive partitioning procedure. Random forest analyses also contain cross-validation routines based on random subsets of survey sites and covariate predictors that are excluded during development of each tree (the 'out-of-bag' data). Cross-validation using out-of-bag data allows estimation of prediction performance (R^2).

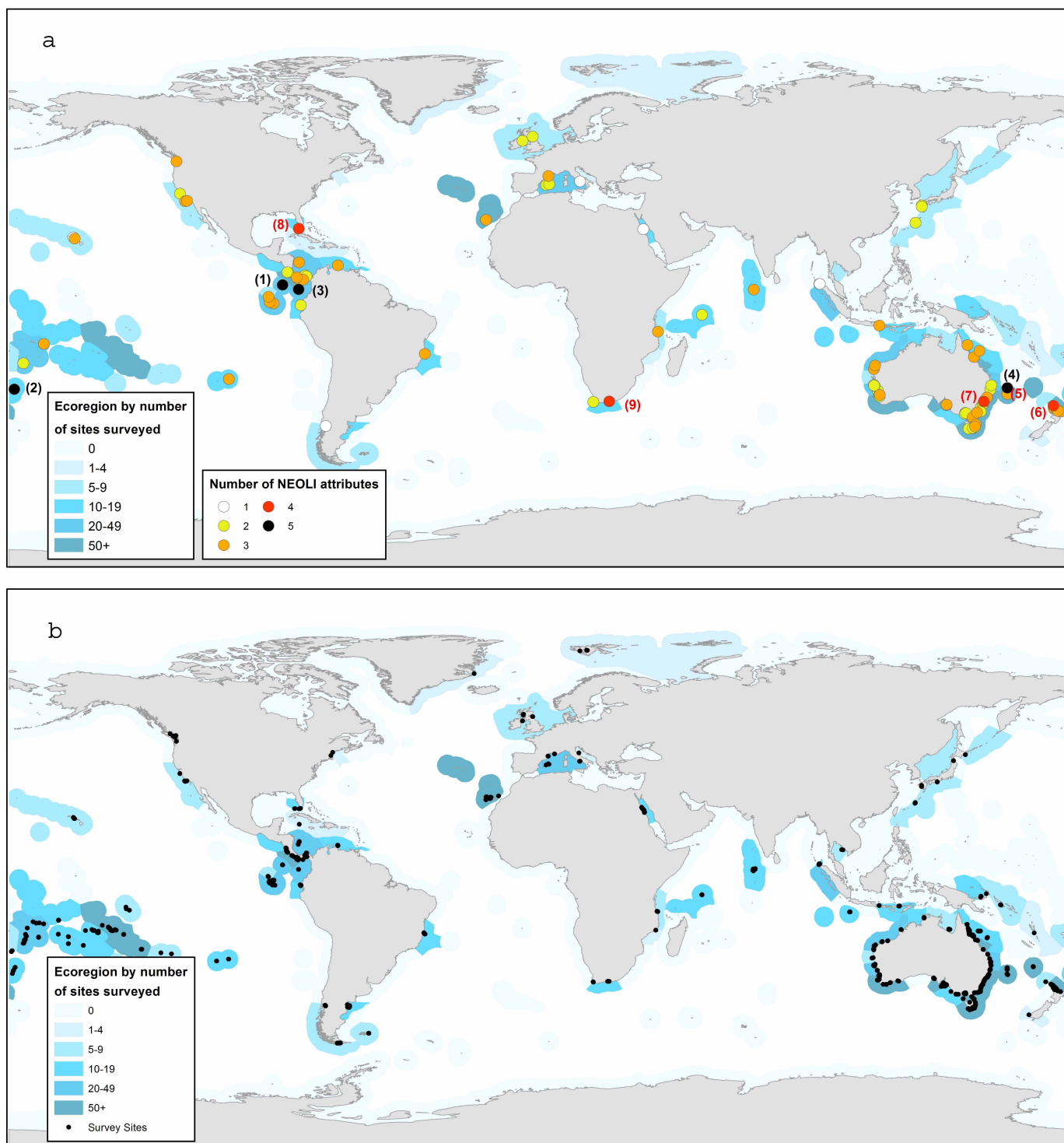
Using random forests, relationships were identified between mean densities of different fish species observed per transect in 76 marine ecoregions¹³ and the global distribution of 14 environmental and socio-economic covariates (Extended Data Table 2). Data for each ecoregion were logged after aggregation as a mean of mean values for sites within each ecoregion, with a total of 1,022 fished sites investigated overall. Ecoregions with a value of zero for a particular metric (for example, grouper biomass in temperate locations) were removed from analysis and treated as missing values when generating predictive models associated with individual MPAs. To estimate prediction error, cross-validation was used where observations not selected in the bootstrap sample for a tree were compared to their predictions. The per cent change in accuracy was measured to assess the importance of each predictor variable (Extended Data Fig. 2). This is the change in accuracy of the predictions between models that include or do not include a given covariate, where accuracy was measured by the mean of the residuals squared using the 'out-of-bag' data.

Linear least-squares regression of survey observations at fished sites with random forest predictions indicated that the models provided a reasonable fit. R^2 values for predicted versus observed plots were 63%, 38%, 80% and 64% for total biomass, large fish biomass, species richness and large fish species richness, respectively, whereas the percentages of observations > predictions were 46%, 46%, 53% and 54%, so observed data were well balanced with an even scatter above and below predictions.

Relationships generated between response metrics and environmental covariates were combined with available data on environmental and socioeconomic covariates at 964 sites surveyed in 87 MPAs to predict each of the eight fish community metrics within each unique combination of MPA zone type (no take or restricted fishing) and ecoregion. From generated random forests, predictions were made at new sites by taking the average of response metrics derived from each tree individually. MPA effects for each MPA zone type were then calculated using the log ratio of predicted/observed value (for example, $\log[B_m/B_p]$), where B_m is measured fish biomass and B_p is biomass predicted if the site was fished). When no individuals of one of the four fish groups (sharks, groupers, jacks or damselfishes) were recorded within a particular MPA, then that MPA was excluded from calculations of effect size. Mean effect sizes and confidence intervals thus relate to the subset of sites where each of the various fish groups were observed.

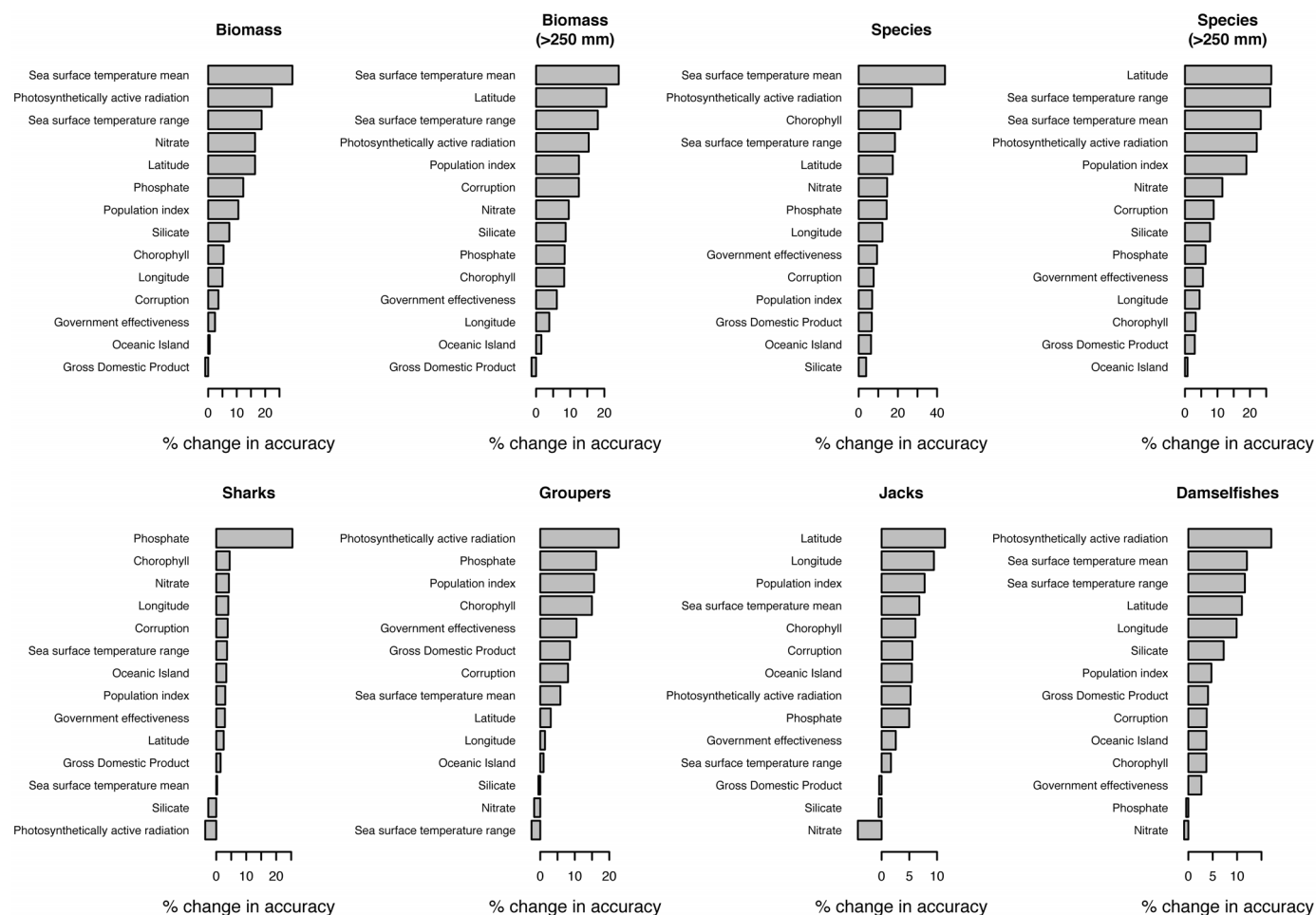
Random forest models were also used to predict values of each metric for fished sites across 5 arcmin grid cells globally, which were then plotted on maps within a coastal buffer. The calculations underlying random forest models used to generate global maps differed from calculations used to predict MPA values in two ways: (1) they were based on 10 rather than 14 environmental covariates, with government effectiveness, corruption, GDP and oceanic island not considered given their small contribution to models (Extended Data Fig. 2) and difficulty in compilation through the full global prediction space; and (2) data for the four fish groups were $\log[x + \text{minimum value for metric}]$ transformed before analysis and back transformed post hoc to compensate for the many zeroes associated with global mapping predictions.

31. Edgar, G. J., Barrett, N. S. & Morton, A. J. Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *J. Exp. Mar. Biol. Ecol.* **308**, 269–290 (2004).
32. CIESIN & CIAT. *Gridded Population of the World, Version 3* (Columbia Univ., 2005).
33. Silverman, B. W. *Density Estimation for Statistics and Data Estimation* (Chapman & Hall, 1986).
34. Kaufmann, D., Kraay, A. & Mastruzzi, M. The Worldwide Governance Indicators: a summary of methodology, data and analytical issues. World Bank Policy Research Working Paper no. 5431, <http://ssrn.com/abstract=1682130> (2010).
35. Tyberghein, L. et al. Bio-ORACLE: a global environmental dataset for marine species distribution modeling. *Glob. Ecol. Biogeogr.* **21**, 272–281 (2012).



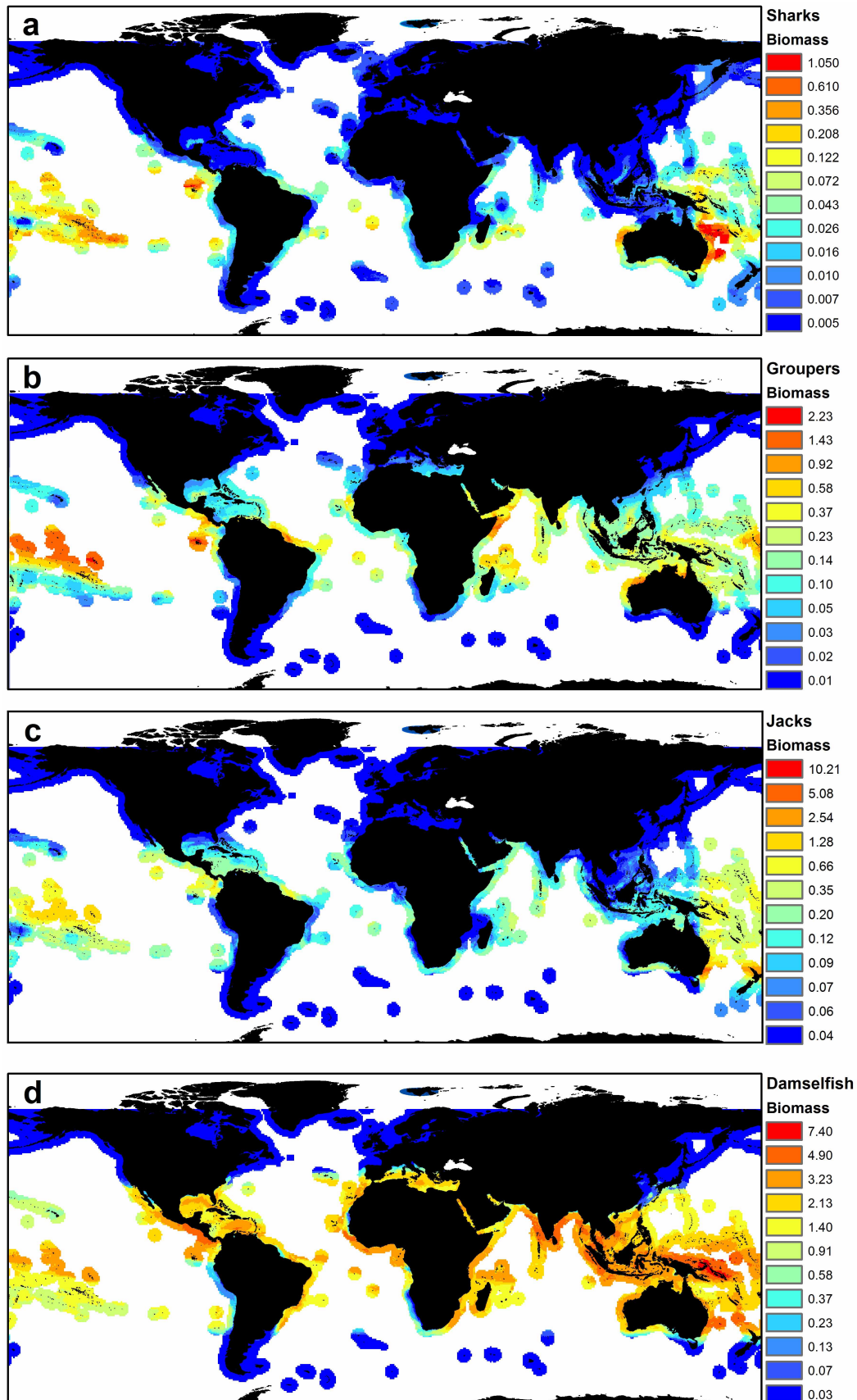
Extended Data Figure 1 | Distribution of sites surveyed. **a**, Number of NEOLI (no take, enforced, old, large and isolated) features at MPAs investigated (coloured circles). MPAs with most NEOLI features are overlaid on top; consequently numerous MPAs with one and two features are not visible. MPAs with five NEOLI features are (1) Cocos, (2) Kermadec Islands,

(3) Malpelo, (4) Middleton Reef; MPAs with four NEOLI features are (5) Elizabeth Reef, (6) Poor Knights Islands, (7) Ship Rock, (8) Tortugas and (9) Tsitsikamma. **b**, All MPA and fished sites surveyed (black circles). Blue shading summarizes the number of sites surveyed within each ecoregion.



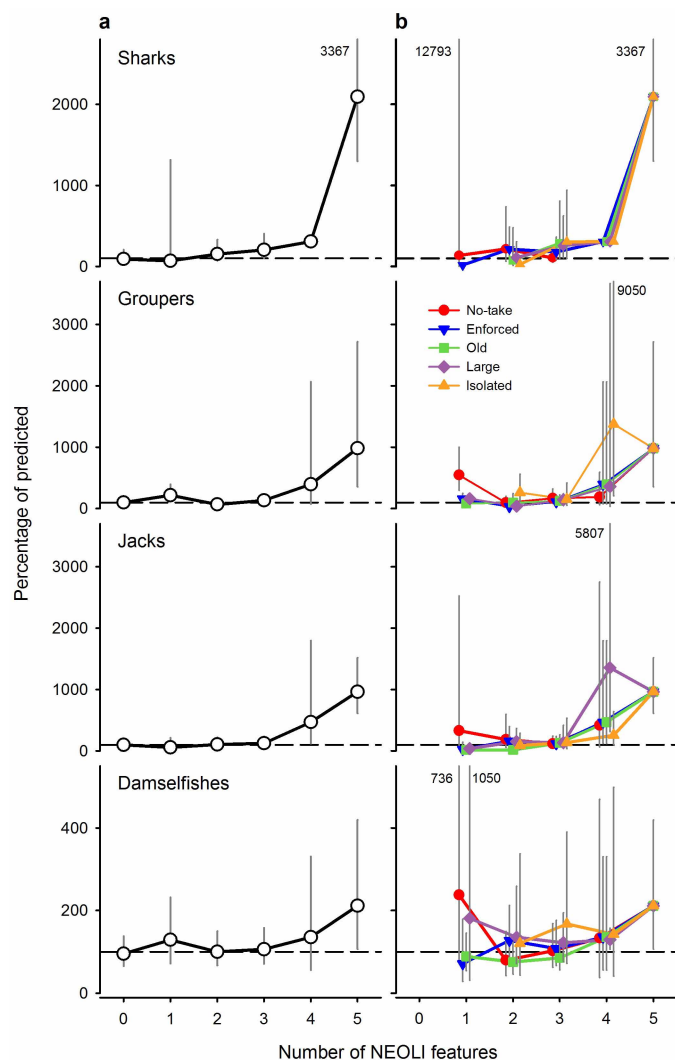
Extended Data Figure 2 | Relative importance of the 14 covariates used in global prediction models developed with random forests. Per cent change in accuracy for a given predictor variable is measured by the change between models that include or do not include that predictor variable, with accuracy

assessed as the mean of the residuals squared. Residuals are based on a cross-validation technique to avoid bias, and the change in accuracy is divided by the standard error for a given tree then averaged across all trees.

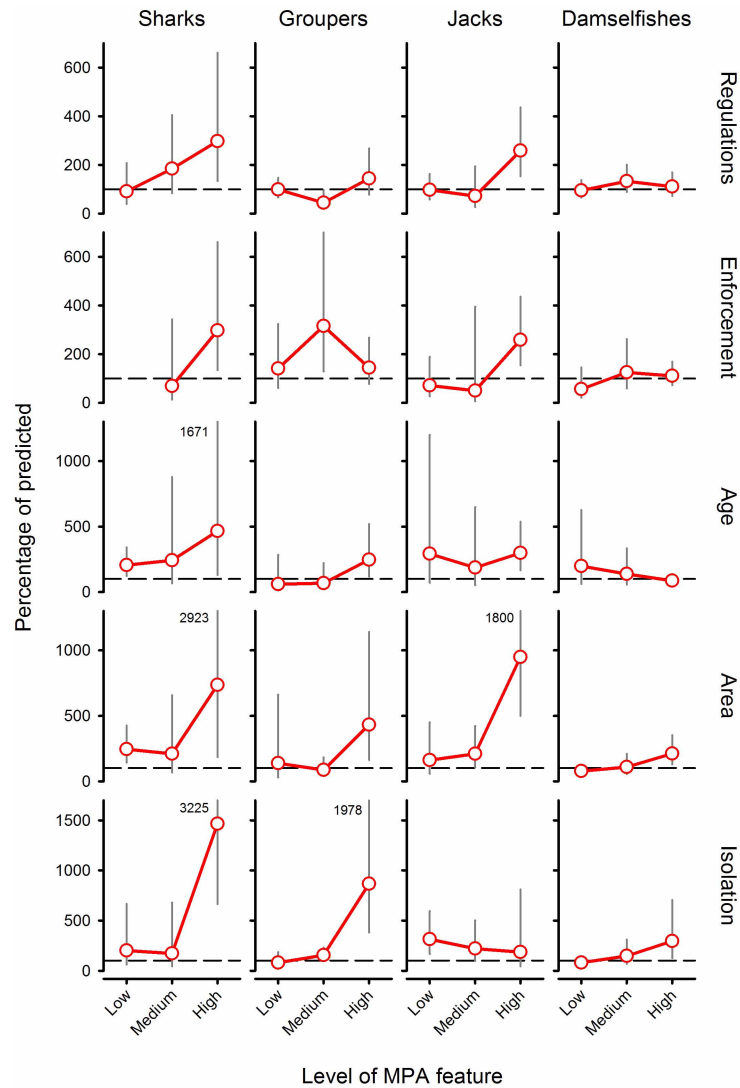


Extended Data Figure 3 | Predicted global distribution of fish biomass (kg per 250 m²) on fished coasts. Predictions are from random forest models developed using data from 1,022 sites in fished locations worldwide. **a**, Sharks.

b, Groupers. **c**, Jacks. **d**, Damselfishes. Note that scales in colour schemes differ among maps, and numbers represent predicted values represented by each colour after smoothing of log-transformed site-level data.



Extended Data Figure 4 | Mean response ratios for MPAs with different number of NEOLI features. Mean ratio values have been back transformed from logs and expressed as percentages with 95% confidence intervals. The number of NEOLI features varies from 0 at sites along fished coastlines to 5 for MPA sites with all NEOLI features. **a**, Plots calculated for sites where sharks, groupers, jacks and damselfishes were present and the subsets of MPAs with different numbers of NEOLI (no take, enforced, old, large, isolated) features. **b**, Mean response ratios for community metrics where each NEOLI feature was included within the set examined. 95% confidence limits that lie off-scale are shown by number. Sample sizes are shown in Extended Data Table 1.



Extended Data Figure 5 | Mean response ratios for the subsets of sites at which sharks, groupers, jacks and damselfishes were observed. Values have been back transformed to per cent, with 100% equivalent to fished coasts, and with 95% confidence intervals. The feature 'regulations' was analysed using data from 82 MPAs that are well enforced; the feature 'enforcement' was analysed using data from 75 MPAs that are no take; and the features 'isolation',

'age' and 'area' were analysed using data from 52 MPAs that are both no take and well enforced. Sharks were not observed in any no-take MPA with low enforcement, so the associated response ratio could not be calculated. 95% confidence limits that lie off-scale are shown by number. Sample sizes are shown in Extended Data Table 1.

Extended Data Table 1 | Sample sizes applied in figures

Figure 2a		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
All metrics	All	76	16	57	39	5	4

Figure 2b		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
All metrics	No-take	7	31	29	4	4	
	Enforced	5	36	32	5	4	
	Old	2	22	31	5	4	
	Large	2	16	16	3	4	
	Isolated		8	9	3	4	

Extended Data Figure 4a		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
Sharks	All	29	3	26	14	1	4
Groupers	All	46	8	36	22	4	4
Jacks	All	50	6	29	27	4	4
Damselfishes	All	62	12	48	36	4	4

Extended Data Figure 4b		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
Sharks	No-take	2	13	9			4
	Enforced	1	22	10	1		4
	Old		5	8	1		4
	Large		9	9	1		4
	Isolated		3	6	1		4
Groupers	No-take	3	19	14	3		4
	Enforced	2	21	16	4		4
	Old	2	14	18	4		4
	Large	1	11	12	3		4
	Isolated		7	6	2		4
Jacks	No-take	2	12	20	3		4
	Enforced	2	21	21	4		4
	Old	1	10	20	4		4
	Large	1	10	11	2		4
	Isolated		4	9	3		4
Damselfishes	No-take	4	25	26	3		4
	Enforced	4	31	29	4		4
	Old	2	17	28	4		4
	Large	2	14	16	2		4
	Isolated		8	9	3		4

Figure 3		Level of feature		
Metric	Feature	Low	Medium	High
All metrics	Regulations	76	30	52
	Enforcement	9	14	52
	Age	5	19	28
	Area	14	30	8
	Isolation	35	9	8

Extended Data Figure 5		Level of feature		
Metric	Feature	Low	Medium	High
Sharks	Regulations	29	15	23
	Enforcement	0	5	23
	Age	3	12	8
	Area	3	14	6
	Isolation	12	6	5
Groupers	Regulations	46	20	27
	Enforcement	7	9	27
	Age	3	8	16
	Area	6	14	7
	Isolation	16	6	5
Jacks	Regulations	50	19	33
	Enforcement	2	6	33
	Age	3	10	20
	Area	8	19	6
	Isolation	19	6	8
Damselfishes	Regulations	62	28	44
	Enforcement	8	10	44
	Age	3	17	24
	Area	11	26	7
	Isolation	29	7	8

Extended Data Table 2 | Covariates used as predictor variables in global random forest models

Variable	Variable abbreviation	Units	Scale	Reference (if applicable)
Index of population pressure	POP_index	index	2.46 arcmin (4.6 km)	
Government Effectiveness	Govt Eff	index	country	35
Control of Corruption	Corruption	index	country	35
Per capita GDP	GDP	US\$	country	#
Mean nitrate	Bio_nitrate	umol/l	5 arcmin (9.2 km)	36
Mean phosphate	BIO_phosphate	umol/l	5 arcmin (9.2 km)	36
Mean silicate	BIO_silicate	umol/l	5 arcmin (9.2 km)	36
Mean chlorophyll A	BIO_chlomean	mg/m ³	5 arcmin (9.2 km)	36
Photosynthetically active radiation	BIO_parmean	Einstein/m ³ /day	5 arcmin (9.2 km)	36
Mean sea surface temperature	BIO_SST_mean	°C	5 arcmin (9.2 km)	36
Range of sea surface temperature	BIO_SST_range	°C	5 arcmin (9.2 km)	36
Oceanic island isolated from continental shelf	Oceanic island	yes/no		
Site latitude	Latitude	decimal degrees	0.0001°	
Site longitude	Longitude	decimal degrees	0.0001°	

The index of population pressure was calculated by fitting a smoothly tapered surface to each settlement point on a year 2000 world population density grid³² using the quadratic kernel function³³. Populations were screened for a density greater than 1,000 people per 0.04 degree cell, and the search radius was set at 3.959 degrees. This table contains refs 34 and 35.

Per capita GDP was obtained from IMF for 2012 at http://en.wikipedia.org/wiki/List_of_countries_by_GDP_%28PPP%29_per_capita.

Global conservation outcomes depend on marine protected areas with five key features

Graham J. Edgar¹, Rick D. Stuart-Smith¹, Trevor J. Willis², Stuart Kininmonth^{1,3}, Susan C. Baker⁴, Stuart Banks⁵, Neville S. Barrett¹, Mikel A. Becerro⁶, Anthony T. F. Bernard⁷, Just Berkhout¹, Colin D. Buxton¹, Stuart J. Campbell⁸, Antonia T. Cooper¹, Marlene Davey¹, Sophie C. Edgar⁹, Günter Försterra¹⁰, David E. Galván¹¹, Alejo J. Irigoyen¹¹, David J. Kushner¹², Rodrigo Moura¹³, P. Ed Parnell¹⁴, Nick T. Shears¹⁵, German Soler¹, Elisabeth M. A. Strain¹⁶ & Russell J. Thomson¹

In line with global targets agreed under the Convention on Biological Diversity, the number of marine protected areas (MPAs) is increasing rapidly, yet socio-economic benefits generated by MPAs remain difficult to predict and under debate^{1,2}. MPAs often fail to reach their full potential as a consequence of factors such as illegal harvesting, regulations that legally allow detrimental harvesting, or emigration of animals outside boundaries because of continuous habitat or inadequate size of reserve^{3–5}. Here we show that the conservation benefits of 87 MPAs investigated worldwide increase exponentially with the accumulation of five key features: no take, well enforced, old (>10 years), large (>100 km²), and isolated by deep water or sand. Using effective MPAs with four or five key features as an unfished standard, comparisons of underwater survey data from effective MPAs with predictions based on survey data from fished coasts indicate that total fish biomass has declined about two-thirds from historical baselines as a result of fishing. Effective MPAs also had twice as many large (>250 mm total length) fish species per transect, five times more large fish biomass, and fourteen times more shark biomass than fished areas. Most (59%) of the MPAs studied had only one or two key features and were not ecologically distinguishable from fished sites. Our results show that global conservation targets based on area alone will not optimize protection of marine biodiversity. More emphasis is needed on better MPA design, durable management and compliance to ensure that MPAs achieve their desired conservation value.

A multitude of socio-economic and biological factors influence the responses of species to protection within MPA networks, adding considerable uncertainty when making specific predictions regarding the conservation benefits of new MPAs. Even within well-designed MPAs, populations of marine species can respond quite differently to prohibitions on fishing as a consequence of species-specific factors such as mobility, larval dispersal, fecundity, longevity, indirect interactions among species, environmental context, and overall level of exploitation before protection^{5,6}. To assess the extent to which MPAs fulfil their ecological potential, we used a database unprecedented in geographic scale to investigate how conservation value, characterized by ecological response of fish communities within MPAs, is affected by the cumulative effects of five key planning and management features: (1) degree of fishing permitted within MPAs; (2) level of enforcement; (3) MPA age; (4) MPA size; and (5) presence of continuous habitat allowing unconstrained movement of fish across MPA boundaries^{6–10}. Although previous studies have considered these factors independently, this is the first study, to

our knowledge, that considers them simultaneously, using data collected globally with standardized methods.

Observations from the subset of MPAs that seem to work effectively—that is, they include at least four of five ‘NEOLI’ (no take, enforced, old, large and isolated) features—are additionally used to infer ecological condition associated with unfished reefs. For this aspect, we used the global network of MPAs as a vast ecological experiment, where effective no-take areas represent human predator exclusion plots within a matrix of fished coasts¹¹.

Eight community-level metrics were assessed using data from 40 nations on shallow reef fish densities and sizes provided by researchers and trained volunteer divers participating in the Reef Life Survey (RLS) programme¹². A total of 964 sites in 87 MPAs were surveyed (Extended Data Fig. 1a), with data aggregated into 121 MPA/ecoregion groupings for analysis. MPA means were compared with statistical predictions for fished coasts using data from 1,022 non-MPA sites surveyed in 76 of the 232 Marine Ecoregions of the World¹³ (Extended Data Fig. 1b and Supplementary Data Table 1). The four community metrics investigated, each widely considered to respond to MPA declaration^{14,15}, were: (1) total biomass of all fishes; (2) total biomass of large (>250 mm length) fishes; (3) species richness of all fishes (number of species sighted per transect); and (4) species richness of large fishes. We also estimated the total biomass of three commercially important taxa (sharks, groupers and jacks), with unexploited damselfishes providing a control group for effects evident on targeted fishery groups. Effect size was calculated using the log ratio of measured values in MPAs relative to values predicted using global models for fished coasts.

Among 14 environmental and socio-economic covariates used in random forest models¹⁶ to develop predictions for fished coasts, mean sea surface temperature, annual temperature range, photosynthetically active radiation, and latitude consistently exerted the strongest influence on the global distribution of species richness and biomass metrics (Extended Data Fig. 2). Biomass of groupers and jacks was also greatly influenced by human population density, and the biomass of sharks and groupers was influenced by phosphate concentration.

Fish species richness along fished coasts peaked in the southeast Asian ‘coral triangle’ region (Fig. 1a), as expected^{12,17}. However, when only the number of large fishes sighted along transects was considered (Fig. 1b), the global centre of species richness shifted to more isolated locations within the Indo-Pacific region. Overfishing of large predatory fishes presumably contributed to these geographical patterns. Sharks, groupers and other large fishes were present within the coral triangle

¹Institute for Marine and Antarctic Studies, University of Tasmania, GPO Box 252-49, Hobart, Tasmania 7001, Australia. ²Institute of Marine Sciences, School of Biological Sciences, University of Portsmouth, Ferry Road, Portsmouth PO4 9LY, UK. ³Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, SE-106 91 Stockholm, Sweden. ⁴School of Plant Science, University of Tasmania, GPO Box 252, Hobart, Tasmania 7001, Australia. ⁵Charles Darwin Foundation, Puerto Ayora, Galapagos, Ecuador. ⁶The Bites Lab, Natural Products and Agrobiological Institute (IPNA-CSIC), 38206 La Laguna, Tenerife, Spain. ⁷Elwandle Node, South African Environmental Observation network, Private Bag 1015, Grahamstown 6140, South Africa. ⁸Wildlife Conservation Society, Indonesia Marine Program, Jalan Atletik No. 8, Bogor Jawa Barat 16151, Indonesia. ⁹Department of Water, Perth, Western Australia 6000, Australia. ¹⁰Facultad de Recursos Naturales, Escuela de Ciencias del Mar, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile. ¹¹Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, Bvd Brown 2915, 9120 Puerto Madryn, Argentina. ¹²Channel Islands National Park, United States National Park Service, 1901 Spinnaker Dr., Ventura, California 93001, USA. ¹³Instituto de Biologia, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Rio de Janeiro 21941-902, Brazil. ¹⁴Scripps Institution of Oceanography, UC San Diego, Mail Code 0227, 9500 Gilman Dr., La Jolla, California 92093-0227, USA. ¹⁵Leigh Marine Laboratory, University of Auckland, 160 Goat Island Road, Leigh 0985, New Zealand. ¹⁶Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, Università di Bologna, Via San Alberto, Ravenna 46100, Italy.

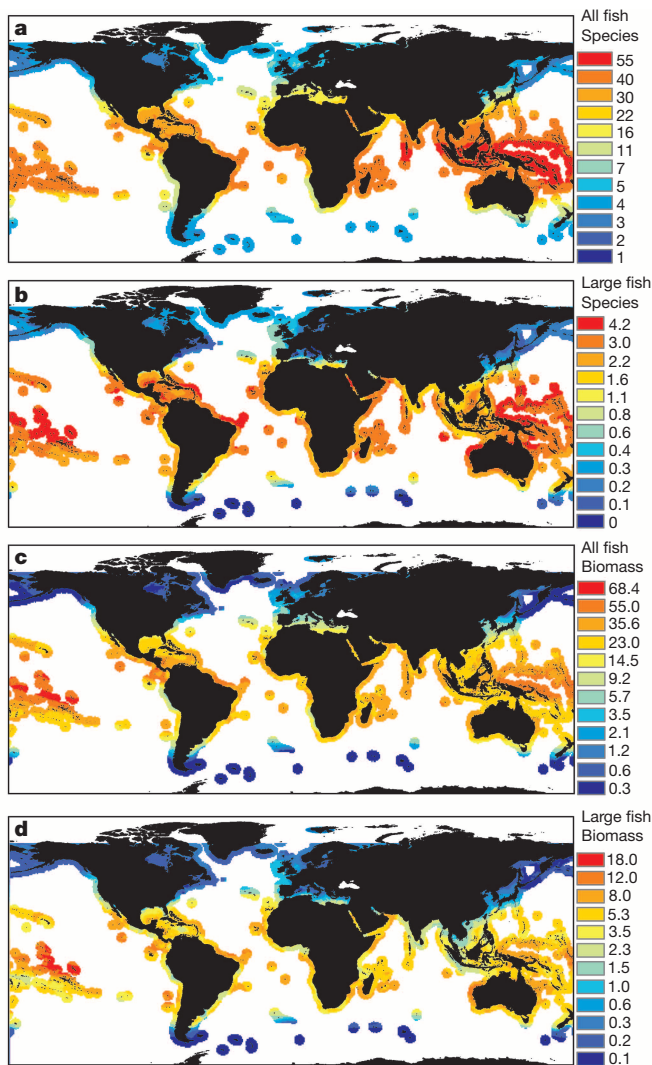


Figure 1 | Predicted global distribution of four community metrics for fishes associated with coral and rocky reefs outside of MPAs. Predictions are from random forest models developed using data from 1,022 sites in fished locations worldwide. **a**, Species richness of all fishes (number of species sighted per 250 m²). **b**, Species richness of large (>250 mm total length) fishes (per 250 m²). **c**, Total biomass of all fishes (kg per 250 m²). **d**, Total biomass of large fishes (kg per 250 m²).

region but had been exploited to near absence on most reefs, and so were rarely recorded on transects; consequently, observed species richness of large fishes was relatively low.

Our predictive models indicated that total fish and large fish biomass were highest in French Polynesia and the nearby Line Islands (Figs 1c, d), and sharks, groupers and jacks also had disproportionally high biomass in that region (Extended Data Figs 3a–c). Shark biomass on fished coasts was also very high off the Pitcairn Island group, and northeastern and northwestern Australia. Reassuringly, high shark and grouper biomass was accurately predicted for Galapagos, regardless that no data from fished sites in the oceanic tropical eastern Pacific region were used to generate the predictive models. At the time of the surveys, all islands in the region (Galapagos, Cocos and Malpelo) were within MPAs; however, data obtained before fishing restrictions in Galapagos indicate anomalously high shark and grouper biomass for fished coasts in that archipelago (S.B. and G.J.E., unpublished data). Damselfishes occurred in relatively high abundance in all tropical ocean basins (Extended Data Fig. 3d).

Across all 87 MPAs investigated, species richness of large fishes was 36% greater inside MPAs compared to fished areas (95% confidence

interval (CI), 16–60% increase), biomass of large fishes was 35% greater (CI 3–78% increase) and sharks 101% greater (CI 17–239% increase). Nevertheless, for species richness of all fishes and the other four biomass metrics investigated, no significant difference ($P > 0.05$) was found between levels observed in MPAs and those predicted for fished coasts. Moreover, many MPAs possessed fish biomass well below predicted regional averages, as indicated by the large percentage of MPAs with negative log ratios for total biomass, ranging from 25% of MPAs for large fishes to 31% for sharks to 47% for groupers. These negative values indicate considerable site-scale variability in fish densities, with some MPA sites exhibiting low fish biomass due to local habitat variability between survey sites and, in other cases, a bias resulting from stakeholder consultation processes before MPA declaration aimed at minimizing lost fishing opportunity¹⁸.

The poor overall performance of MPAs worldwide in terms of recovery of fish biomass relative to fished sites was due to a high frequency of ineffective MPAs and high spatial variability in fish densities, rather than an absence of recovery in all MPAs. The efficacy of MPAs was strongly influenced by the five NEOLI planning and management features (no take, enforced, old, large and isolated), with MPAs that scored highly with multiple NEOLI features typically having highly elevated biomass of exploitable fishes compared to fished sites (Fig. 2). MPAs with at least four NEOLI features were distributed across six countries in three oceans (Extended Data Fig. 1a) and a range of environmental conditions, indicating that model outputs and conclusions were not strongly regionally biased.

No significant differences were evident between fished sites (zero features) and MPAs with one or two NEOLI features; however, effect sizes rose rapidly when the number of features increased from three to five (Fig. 2 and Extended Data Fig. 4). For example, the measured rises in mean values within MPAs relative to fished areas for total fish biomass, total large fish biomass and shark biomass with three NEOLI features were 30%, 66% and 104%, respectively. These increases were, however, modest compared to values when all five NEOLI features were present, with large increases of 244%, 840% and 1,990%, respectively. Similar marked increases in biomass were evident for groupers (882%) and jacks (864%). Non-fished damselfishes showed a smaller mean increase of 111% at MPAs with five NEOLI features. This increase was on the margins of statistical significance, lying outside the 95% confidence interval (Extended Data Fig. 4) but nonsignificant ($P < 0.05$) when assessed with a *t*-test, which adjusts for small sample size.

All four MPAs with five NEOLI features were small oceanic islands (Cocos, Costa Rica; Malpelo, Colombia; Kermadec Islands, New Zealand; and Middleton Reef, Australia), raising a potential concern that calculated effect sizes were biased by plankton and pelagic fish subsidies that enlarge food webs at isolated oceanic locations. ‘Oceanic island’ was, however, included as a categorical covariate in random forest models, therefore model predictions should accommodate small island effects. Regardless, further investigation into the contribution of external subsidies to food webs at isolated MPAs is warranted. Alternative explanations for elevated damselfish numbers in the most effective MPAs compared with poorly protected MPAs include reduced fishing-related habitat deterioration such as dynamite damage to coral, and trophic cascades involving smaller predators that consume damselfishes and are prey to sharks and groupers.

No-take regulations, efficient enforcement, large area (>100 km²) and old age (>10 years) each contributed similar increases in fish biomass within MPAs (Fig. 2). However, isolation, a categorical factor that distinguished MPAs with reef habitat surrounded by deep (>25 m) water or large expanses of sand from MPAs with shallow reef habitat extending to fished areas, seemed to exert a stronger influence for community-level biomass and richness metrics than the other four features. For example, the mean increase (95% CI) for total fish biomass associated with MPAs with three NEOLI features was 100% (14–252%) when one of the three features was isolation, compared to 14% (–18%–58%) for three NEOLI MPAs when isolation was not included. Compliance

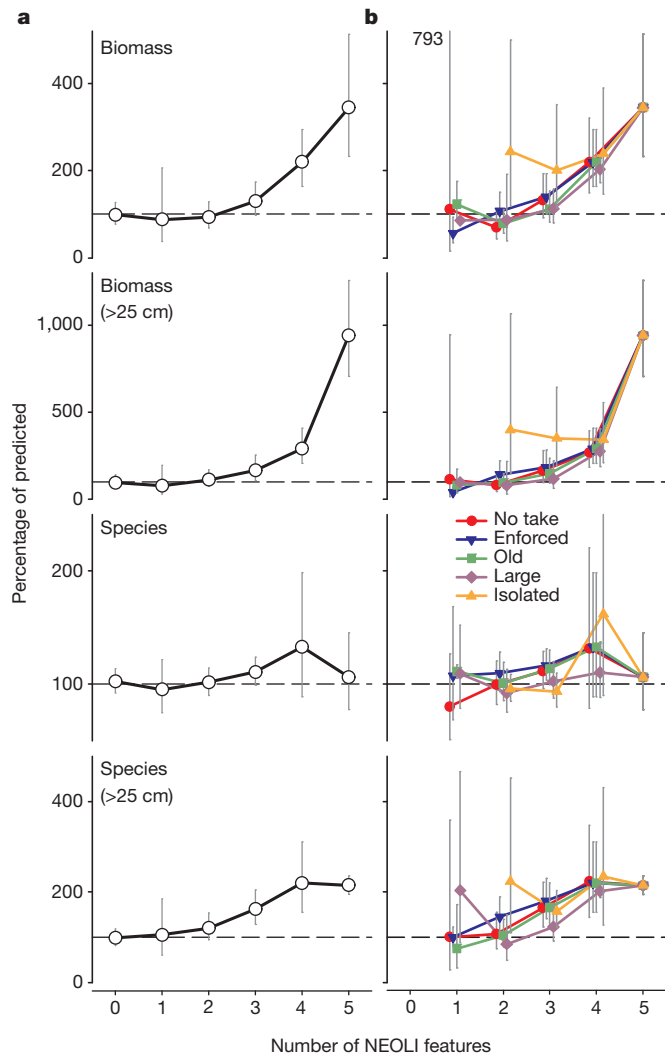


Figure 2 | Mean response ratios for MPAs with different numbers of NEOLI (no take, enforced, old, large, isolated) features. Mean ratio values have been back transformed from logs and expressed as percentages with 95% confidence intervals, with 100% equivalent to fished coasts. Sites on fished coasts have 0 NEOLI features. **a**, Mean response ratios for four community metrics. **b**, Mean response ratios for community metrics where each NEOLI feature was included within the set examined. The 'no-take' plot with two features, for example, depicts the mean response for no-take MPAs with a single other NEOLI feature. 95% confidence limits that lie off-scale are shown by number. Samples sizes are shown in Extended Data Table 1.

may have contributed to the isolation effect, in that isolated MPAs are generally well demarcated for control purposes. They are readily recognized by fishers and more easily policed than coastlines with complicated mosaics of no take, restricted take and fishing zones. Although very important, the effect of isolation was similar in magnitude—rather than clearly superior—to other MPA features for biomass of sharks, groupers and jacks (Extended Data Fig. 4).

When MPAs that are no take and well enforced are considered, differences were evident in how the other MPA features affect different components of the fish community (Fig. 3 and Extended Data Fig. 5). Total fish biomass increased significantly from low to high levels for all five MPA features, and these same trends were magnified for large fishes (Fig. 3). Regardless of general concerns that large pelagic species move such great distances that few individuals are fully protected within MPAs¹⁹, sharks and jacks seem to receive considerable protection from fishing mortality within the large, well-enforced, no-take MPAs studied here. The biomass of sharks and groupers rose exponentially when MPAs were fully isolated, and also greatly increased with area and age. The

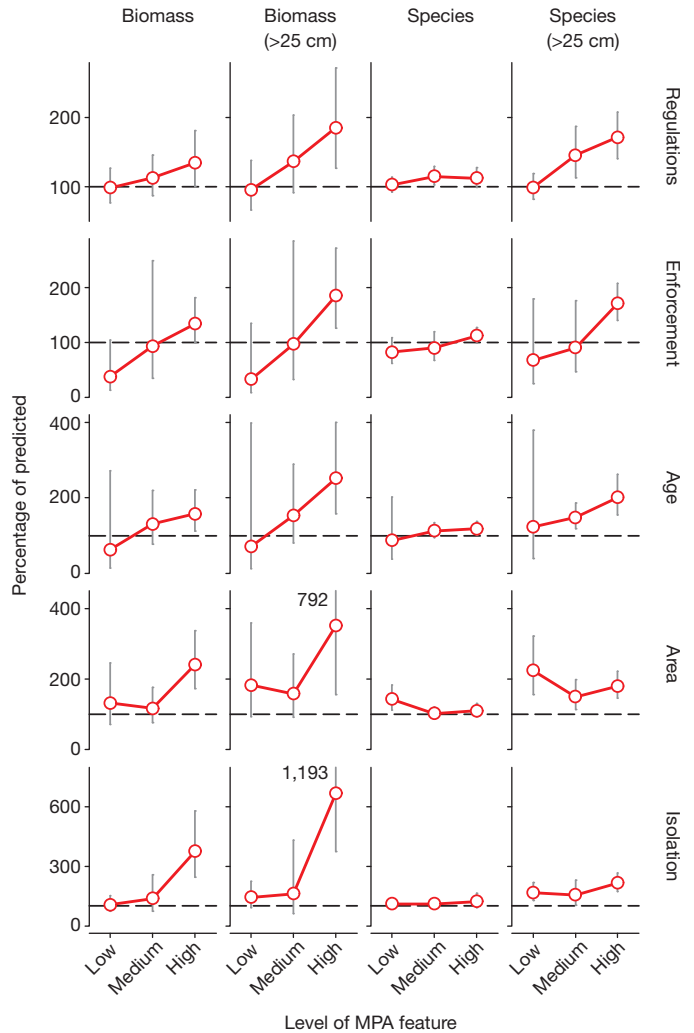


Figure 3 | Mean response ratios with 95% confidence intervals for four community metrics and low, medium and high levels of five MPA features. Values have been back transformed to per cent scale, with 100% equivalent to fished coasts. The feature 'regulations' was analysed using data from 82 MPAs with high enforcement; the feature 'enforcement' was analysed using data from 75 MPAs that are no-take; and the features 'isolation', 'age' and 'area' were analysed using data from 52 MPAs that are both no take and well enforced. 95% confidence limits that lie off-scale are shown by number. Samples sizes are shown in Extended Data Table 1.

biomass of jacks showed little isolation and age effects, but rose greatly in MPAs that were large, well enforced and no take. Damselfish biomass did not increase significantly with the accumulation of individual NEOLI features.

The large number of MPAs investigated here has allowed relatively subtle and higher order interactive MPA effects to be detected. Previous studies of MPAs have shown, for example, negligible or weak patterns associated with MPA size^{6,9,14,15,20}, and those detected here were only evident for MPAs with at least three of the NEOLI features. However, MPA size was very important for such metrics as jack biomass, which showed a stronger response to MPA area than to other metrics (Extended Data Fig. 5). This response probably resulted from time spent by actively-swimming fishes outside park boundaries, which increases probability of capture for fishes associated with small MPAs.

Species richness of large fishes exhibited a highly significant difference between MPAs with five NEOLI features and fished locations (115% increase relative to predicted, CI 95–137%; *t*-test, $P < 0.0001$; Fig. 2). By contrast, MPAs with five NEOLI features did not differ significantly in total species richness (6% increase relative to predicted)

from fished locations (t -test, $P = 0.42$; Fig. 2), nor did any of the five features individually have a clear effect on species richness (Fig. 3). Thus, total species richness along transects did not detectably increase in effective MPAs, despite the presence of additional large fish species, perhaps because of food web changes in the form of reduced presence of small fish species that comprised prey of the larger predatory species^{5,21,22}. Regardless of these transect-scale effects, species richness at regional scales probably increased in areas with a mosaic of fished and effective MPAs because of the additional presence of large fishery-targeted species within the seascape¹⁸.

Of the 87 MPAs investigated, only four possessed all five NEOLI features, whereas five MPAs possessed four features, and 39, 57 and 16 MPAs possessed three, two and one feature, respectively. The low proportion of MPAs possessing four or five NEOLI features (10%), and thus regarded here as effective, probably overstates the true proportion of effective MPAs worldwide. Our survey strategy deliberately targeted well-known and well-regarded MPAs, with most large and long-established MPAs included in this study.

Although only a small subset of MPAs are ever likely to qualify as large, most MPAs could achieve the remaining four NEOLI features. MPAs require additional time to age, and sufficient will among stakeholders, managers and politicians for increased implementation of no-fishing zones, increased levels of compliance, and extension of boundaries past the limits of reef systems or to deep water. If these could be achieved in tandem with current trends for declaration of large remote 'wilderness' MPAs^{23,24}, then conservation benefits from the global MPA network should increase markedly. However, the current base is very low with only 0.08% of the world's oceans within no-take MPAs in 2008 (ref. 25), and with opportunities for an expanded network diminishing as establishment and opportunity costs for large isolated MPAs escalate in line with human population growth^{24,26}.

By using effective MPAs as an unfished standard, our study allows the first global assessment of the magnitude of fishing effects on temperate as well as tropical reef communities. Fish biomass was greatly reduced overall, with 63% of all fish biomass, 80% of large fish biomass, 93% of sharks, 84% of groupers and 85% of jacks apparently removed from reefs by fishing.

In spite of their huge magnitude, these estimates are probably conservative because they are based on the assumption that MPAs with four or five NEOLI features provide an accurate non-fished baseline for inferring historical patterns. Yet fish populations are unlikely to have fully recovered from previous impacts of fishing in four NEOLI MPAs, which were found to be less effective than five NEOLI MPAs for some metrics. Moreover, high fishing mortality rates for sharks and wide-ranging predatory fishes outside MPAs will negatively influence total numbers within boundaries through reduced immigration rates, and further recovery of fish biomass within MPAs probably continues over much longer time spans than the 10-year threshold used here to define old MPAs¹⁸. Our estimates for effective MPAs include uncertainty associated with the low number of effective MPAs surveyed, most notably for sharks, as only five of the nine category 4 and 5 NEOLI MPAs had sharks present. Also, biomass may be overestimated because of diminished flee responses from divers of large fishes in well-enforced no-take MPAs²⁷. Regardless, fishing clearly exerts a very large and ubiquitous impact on shallow reefs.

The 80% reduction in biomass of large fishes outside effective MPAs coincides with the threshold value used by the International Union for Conservation of Nature (IUCN) to categorize species as Critically Endangered for Red List assessments²⁸. Although recognizing that application of current Red List thresholds to exploited fish stocks remains contentious²⁸, the high number of large-bodied species that together average 80% decline indicates that innumerable threatened fish species probably exist, and that effective MPAs probably have a large role in safeguarding populations of many of these species⁴. Even nations with relatively well-managed fisheries have few sharks and other large predatory coastal fishes outside well-designed and mature MPAs. Given

the huge scale of fishing impacts, the rate of fish extinctions is likely to increase greatly through this century unless a refugial network of effective MPAs exists to allow persistence of large-bodied species and associated predator-dominated food webs, and broad-scale fisheries management practices significantly improve²⁹.

METHODS SUMMARY

Surveys were based on Reef Life Survey methodology^{12,30}, with support from volunteer SCUBA divers trained individually to scientific data collection standards. All fishes observed within 50 m × 5 m transect blocks were counted, and total fish lengths estimated, during swims on adjoining blocks up one side and down the other side of 50-m lines. Each transect was set along a depth contour, with two depth contours (mean 2.4) generally surveyed at each site. Sites located within 87 MPAs were investigated, with approximately half located in Australia (36) and New Zealand (8). In total, 48 MPAs were complete no take, 18 MPAs allowed limited fishing, whereas 21 MPAs were multi-zoned with interspersed no-take and limited fishing zones. Data were compiled from 171,331 underwater abundance counts of 2,544 species in 9,544 transect blocks at 1,986 sites.

We assessed effects of five MPA features (fishing regulations, enforcement, age, area and isolation), each categorized at low, medium and high levels, on eight fish community metrics (species richness of all fishes and large (>250 mm) fishes; total biomass of all fishes, large fishes, sharks, groupers, jacks and damselfishes). The magnitudes of effects were quantified using the log ratio of observed value within the MPA to predicted value at that location if the MPA did not exist (for example, $\log[B_m/B_p]$, where B_m is measured fish biomass and B_p is biomass predicted if the site was fished). Predictions were produced using random forest procedures¹⁶, where each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. Relationships were initially established between 14 covariates (environmental and socio-economic) and measured values of the eight response metrics at fished sites. These relationships were then used, with known covariate values at each MPA, to predict each of the eight community metrics at that MPA location.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 September 2013; accepted 13 January 2014.

Published online 5 February; corrected online 12 February 2014 (see full-text HTML version for details).

- Hilborn, R., Micheli, F. & De Leo, G. A. Integrating marine protected areas with catch regulation. *Can. J. Fish. Aquat. Sci.* **63**, 642–649 (2006).
- Gell, F. R. & Roberts, C. M. Benefits beyond boundaries: the fishery effects of marine reserves. *Trends Ecol. Evol.* **18**, 448–455 (2003).
- Edgar, G. J. Does the global network of marine protected areas provide an adequate safety net for marine biodiversity? *Aquat. Conserv.* **21**, 313–316 (2011).
- Mora, C. et al. Coral reefs and the global network of Marine Protected Areas. *Science* **312**, 1750–1751 (2006).
- Babcock, R. C. et al. Decadal trends in marine reserves reveal differential rates of change in direct and indirect effects. *Proc. Natl Acad. Sci. USA* **107**, 18256–18261 (2010).
- Claudet, J. et al. Marine reserves: size and age do matter. *Ecol. Lett.* **11**, 481–489 (2008).
- Guidetti, P. et al. Italian marine reserve effectiveness: does enforcement matter? *Biol. Conserv.* **141**, 699–709 (2008).
- Le Quesne, W. J. F. Are flawed MPAs any good or just a new way of making old mistakes? *ICES J. Mar. Sci.* **66**, 132–136 (2009).
- Vandeperre, F. et al. Effects of no-take area size and age of marine protected areas on fisheries yields: a meta-analytical approach. *Fish. Fish.* **12**, 412–426 (2011).
- Forcada, A., Bayle-Sempere, J. T., Valle, C. & Sanchez-Jerez, P. Habitat continuity effects on gradients of fish biomass across marine protected area boundaries. *Mar. Environ. Res.* **66**, 536–547 (2008).
- Walters, C. J. & Holling, C. S. Large-scale management experiments and learning by doing. *Ecology* **71**, 2060–2068 (1990).
- Stuart-Smith, R. D. et al. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539–542 (2013).
- Spalding, M. D. et al. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
- Côté, I. M., Mosquera, I. & Reynolds, J. D. Effects of marine reserve characteristics on the protection of fish populations: a meta-analysis. *J. Fish. Biol.* **59** (Suppl. A), 178–189 (2001).
- Halpern, B. S. The impact of marine reserves: do reserves work and does reserve size matter? *Ecol. Appl.* **13** (Suppl.), 117–137 (2003).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 15–32 (2001).
- Roberts, C. M. et al. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
- Edgar, G. J., Barrett, N. S. & Stuart-Smith, R. D. Exploited reefs protected from fishing transform over decades into conservation features otherwise absent from seascapes. *Ecol. Appl.* **19**, 1967–1974 (2009).

19. Grüss, A., Kaplan, D. M., Guénette, S., Roberts, C. M. & Botsford, L. W. Consequences of adult and juvenile movement for marine protected areas. *Biol. Conserv.* **144**, 692–702 (2011).
20. Guidetti, P. & Sala, E. Community-wide effects of marine reserves in the Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **335**, 43–56 (2007).
21. Graham, N. A. J., Evans, R. D. & Russ, G. R. The effects of marine reserve protection on the trophic relationships of reef fishes on the Great Barrier Reef. *Environ. Conserv.* **30**, 200–208 (2003).
22. Willis, T. J. & Anderson, M. J. Structure of cryptic reef fish assemblages: relationships with habitat characteristics and predator density. *Mar. Ecol. Prog. Ser.* **257**, 209–221 (2003).
23. Pala, C. Giant marine reserves pose vast challenges. *Science* **339**, 640–641 (2013).
24. Graham, N. A. J. & McClanahan, T. R. The last call for marine wilderness? *Bioscience* **63**, 397–402 (2013).
25. Wood, L. J., Fish, L., Laughren, J. & Pauly, D. Assessing progress towards global marine protection targets: shortfalls in information and action. *Oryx* **42**, 1–2 (2008).
26. Mora, C. *et al.* Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *PLoS Biol.* **9**, e1000606 (2011).
27. Cole, R. G. Abundance, size structure, and diver-oriented behaviour of three large benthic carnivorous fishes in a marine reserve in northeastern New Zealand. *Biol. Conserv.* **70**, 93–99 (1994).
28. IUCN Standards and Petitions Working Group. *Guidelines for Using the IUCN Red List Categories and Criteria* Version 7.0 (downloaded 18 November 2013 from <http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf>, 2008).
29. Mora, C. & Sale, P. F. Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Mar. Ecol. Prog. Ser.* **434**, 251–266 (2011).
30. Edgar, G. J. & Stuart-Smith, R. D. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Mar. Ecol. Prog. Ser.* **388**, 51–62 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the many Reef Life Survey (RLS) divers who contributed to data collection. Development of the RLS data set was supported by the former Commonwealth Environment Research Facilities Program, whereas analyses were supported by the Australian Research Council, a Fulbright Visiting Scholarship (to G.J.E.), the Institute for Marine and Antarctic Studies, and the Marine Biodiversity Hub, a collaborative partnership funded under the Australian Government's National Environmental Research Program. Surveys were assisted by grants from the National Geographic Society, Conservation International, Wildlife Conservation Society, Winifred Violet Scott Trust, Tasmanian Parks and Wildlife Service, the Winston Churchill Memorial Trust, University of Tasmania, and ASSEMBLE Marine. We are grateful to the many park officers who assisted the study by providing permits and assisting with field activities, and to numerous marine institutions worldwide for hosting survey trips.

Author Contributions G.J.E. and R.D.S.-S. conceived the project; G.J.E., R.D.S.-S., M.A.B., A.T.F.B., S.C.B., S.B., S.J.C., A.T.C., M.D., S.C.E., G.F., D.E.G., A.J.I., S.K., D.J.K., R.M., G.S., E.M.A.S. and many others collected the data; G.J.E., R.J.T., T.J.W., S.K. and S.C.E. prepared figures; G.J.E. drafted the initial manuscript; all authors contributed to analyses and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.E. (g.edgar@utas.edu.au).

METHODS

Survey methodology. Standardized visual census counts were made at 1,986 sites using Reef Life Survey (RLS) methodology (see RLS methods manual 'Standardised survey procedures for monitoring rocky and coral reef ecological communities' at http://reeflifesurvey.com/files/2008/09/NEW-Methods-Manual_15042013.pdf). Divers made counts and estimates of total lengths of all fish species observed during swims at $\sim 2 \text{ m s}^{-1}$ along the centre of a 5-m-wide swathe up one side and then down the other side of 50-m transect lines. Fishes sighted in transect blocks were recorded on an underwater slate, with abundance estimates made by counting individuals of less abundant species and, in locations with high fish densities, estimating the number of more abundant species. The abundance of schooling fishes was recorded by counting a subset within the school which was combined with an estimate of the proportion of the total school. Nearly all fishes were recorded to species level, with exceptions classified at the highest taxonomic resolution possible. The use of digital photography typically allowed later identification of most unidentified species, with assistance of taxonomic experts as required.

Experienced scientists and skilled recreational divers both contributed data to the RLS programme, all divers having either substantial previous experience in fish surveys or extensive one-on-one training by R.D.S.-S. or G.J.E. To provide a major element of consistency in diver contributions at the global scale, G.J.E. and R.D.S.-S. participated in most surveys, providing 31% of all data analysed. Validation tests indicated no difference in quality or composition of data provided by volunteers participating in this programme when compared to professional biologists³⁰.

Each transect was set along a depth contour, with two depth contours generally surveyed at each site (mean of 2.4 depths per site; minimum, maximum, mean \pm s.d. depth contours surveyed: 0.1 m, 42 m, 7.5 ± 4.1 m, respectively). Sites located within 87 MPAs were investigated, with approximately half located in Australia (36) and New Zealand (8). In total, 48 MPAs were no take where all fishing was prohibited, 18 MPAs allowed limited fishing, whereas 21 MPAs were multi-zoned with interspersed no-take and limited fishing zones. Data were compiled from 171,331 underwater abundance counts of 2,544 species in 9,544 transect blocks (50 m \times 5 m).

MPA features and community metrics investigated. We assessed the influence of five MPA features on eight fish community metrics calculated using field survey data. The MPA features investigated were each categorized at three levels: low (L), medium (M) and high (H). (1) Regulations. Extent that regulations restrict fishing at survey site. L, site can be openly fished with no fishing restrictions additional to those generally applied within the state; M, site located within an MPA but with some fishing methods allowed; H, no-take area within an MPA. (2) Enforcement. Extent of compliance to regulations that restrict fishing, both through overt policing and through community support for regulations. Level was decided at the time of surveys after discussion with local park authorities, and on the basis of observations of the extent of infractions while conducting fieldwork. L, little attempt at control, a 'paper park'; M, a moderate level of policing attempted, although infractions were apparent; H, appears to be well enforced, although clandestine poaching may occur. (3) Age. Period between when regulations restricting fishing were first enacted and field surveys undertaken. L, MPA zone <5 years old; M, MPA zone 5–10 years old; H, MPA zone >10 years old. (4) Area. MPA zone area, as described in management plan or documents provided locally to users. L, <1 km²; M, 1–100 km²; H, >100 km². (5) Isolation. Degree that reef habitat surveyed is isolated by habitat boundaries from adjacent fished reef. L, shallow (<25 m) reef habitat extends continuously across MPA boundary; M, a small (1–20%) percentage of zone boundary breached by continuous shallow reef habitat; H, MPA zone isolated from fishing areas by depth (>25 m) or sand barriers of at least 20 m width.

We investigated eight community metrics. (1) Species richness of all fishes. Total number of all fish species sighted within 50 m \times 5 m transect blocks. (2) Species richness of large fishes. Total number of fish species sighted within 50 m \times 5 m transect blocks for the set of individuals observed on transects exceeding the 250 mm size class bin (that is, 300 mm size and above). (3) Total fish biomass. Total biomass of all fishes sighted in 50 \times 5 m transect blocks. Estimated by combining abundance counts with size estimates using length-weight relationships provided for total length of each fish species (in some cases genus and family) in Fishbase (<http://www.fishbase.org>). Bias in divers' perceptions of fish size underwater was additionally corrected using relationships presented in ref. 31. (4) Total biomass of large fishes. Total biomass of individuals sighted in 50 \times 5 m transect blocks that exceeded the 250 mm size class bin. (5) Total biomass of sharks. Sum of biomass of all fishes in transect that belong to orders Carcharhiniformes, Heterodontiformes, Lamniformes and Orectolobiformes. (6) Total biomass of groupers. Sum of biomass of all fishes in transect that belong to family Serranidae, genera *Dermatolepis*, *Epinephelus*, *Gracila*, *Mycteroperca*, *Paralabrax*, *Plectropomus*, *Trachypoma* and *Variola*. Small serranids such as *Pseudanthias* spp. were not considered. (7) Total biomass of jacks. Sum of biomass of all fishes in transect that belong to family Carangidae. (8) Total biomass of damselfishes. Sum of biomass of all fishes in transect that belong to family Pomacentridae.

Data aggregation. To reduce spatial confounding resulting from highly clumped distribution of sites surveyed, data were aggregated before analyses as means for each ecoregion, MPA and zone type. Thus, fished sites were aggregated as mean values for each of 76 Marine Ecoregions of the World¹³, whereas MPA data were aggregated into 121 MPA zones by ecoregion combinations. Multi-zoned MPAs contributed two data points to analyses (no-take sites and restricted fishing sites), whereas very large MPAs that extended across ecoregional boundaries (for example, Great Barrier Reef Marine Park, Galapagos Marine Reserve) were also partitioned with aggregated data from each ecoregion.

Global models. Models were developed using random forest procedures¹⁶, as available in the 'extendedForest' packages for R (<https://r-forge.r-project.org/projects/gradientforest/>), to predict the distribution of the eight community metrics in inshore habitats globally, including the MPA locations investigated. Each random forest consisted of numerous (2,000 in this case) regression trees, where each tree is fit to a bootstrap sample of the biological data using a recursive partitioning procedure. Random forest analyses also contain cross-validation routines based on random subsets of survey sites and covariate predictors that are excluded during development of each tree (the 'out-of-bag' data). Cross-validation using out-of-bag data allows estimation of prediction performance (R^2).

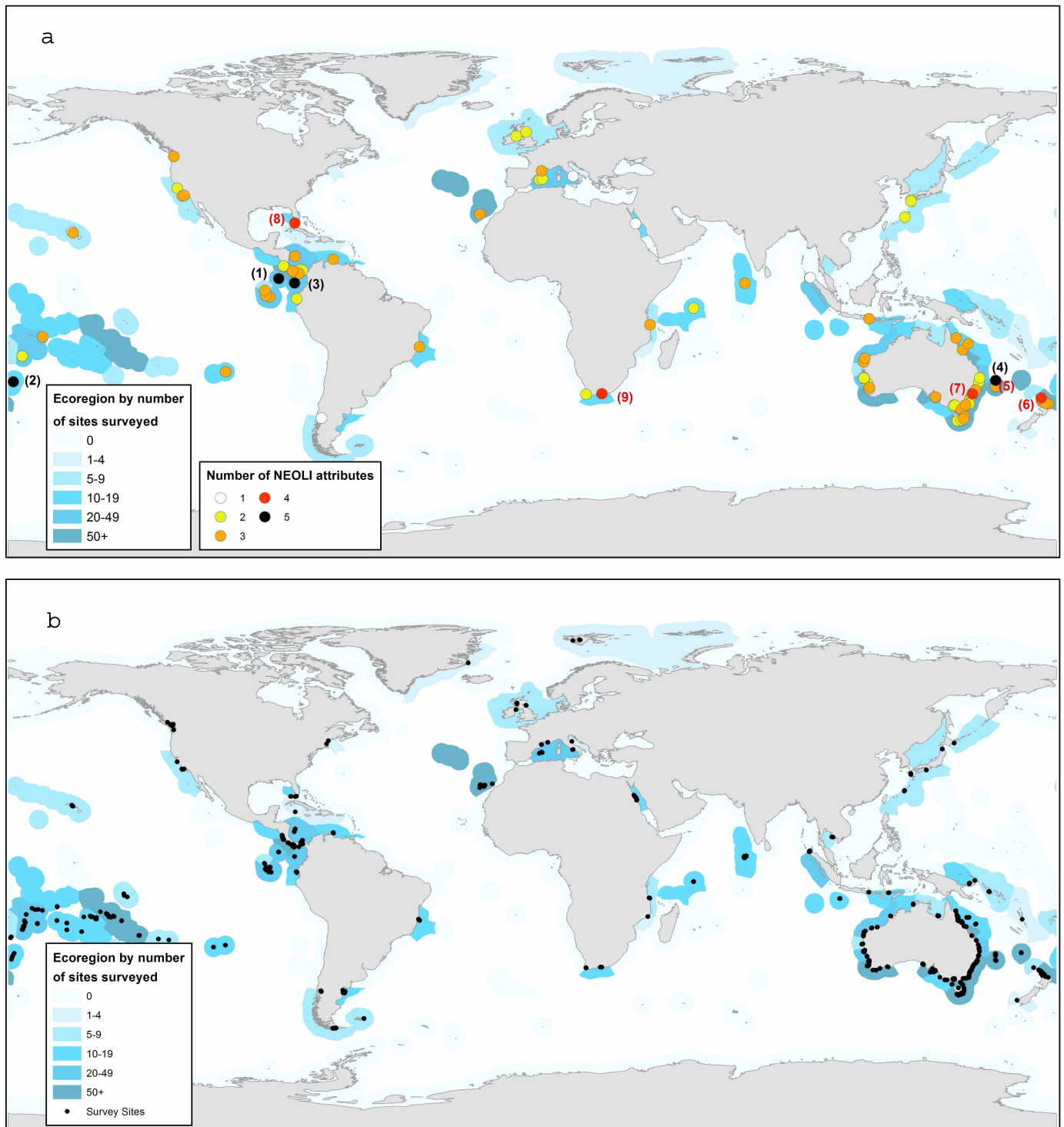
Using random forests, relationships were identified between mean densities of different fish species observed per transect in 76 marine ecoregions¹³ and the global distribution of 14 environmental and socio-economic covariates (Extended Data Table 2). Data for each ecoregion were logged after aggregation as a mean of mean values for sites within each ecoregion, with a total of 1,022 fished sites investigated overall. Ecoregions with a value of zero for a particular metric (for example, grouper biomass in temperate locations) were removed from analysis and treated as missing values when generating predictive models associated with individual MPAs. To estimate prediction error, cross-validation was used where observations not selected in the bootstrap sample for a tree were compared to their predictions. The per cent change in accuracy was measured to assess the importance of each predictor variable (Extended Data Fig. 2). This is the change in accuracy of the predictions between models that include or do not include a given covariate, where accuracy was measured by the mean of the residuals squared using the 'out-of-bag' data.

Linear least-squares regression of survey observations at fished sites with random forest predictions indicated that the models provided a reasonable fit. R^2 values for predicted versus observed plots were 63%, 38%, 80% and 64% for total biomass, large fish biomass, species richness and large fish species richness, respectively, whereas the percentages of observations > predictions were 46%, 46%, 53% and 54%, so observed data were well balanced with an even scatter above and below predictions.

Relationships generated between response metrics and environmental covariates were combined with available data on environmental and socioeconomic covariates at 964 sites surveyed in 87 MPAs to predict each of the eight fish community metrics within each unique combination of MPA zone type (no take or restricted fishing) and ecoregion. From generated random forests, predictions were made at new sites by taking the average of response metrics derived from each tree individually. MPA effects for each MPA zone type were then calculated using the log ratio of predicted/observed value (for example, $\log[B_m/B_p]$), where B_m is measured fish biomass and B_p is biomass predicted if the site was fished). When no individuals of one of the four fish groups (sharks, groupers, jacks or damselfishes) were recorded within a particular MPA, then that MPA was excluded from calculations of effect size. Mean effect sizes and confidence intervals thus relate to the subset of sites where each of the various fish groups were observed.

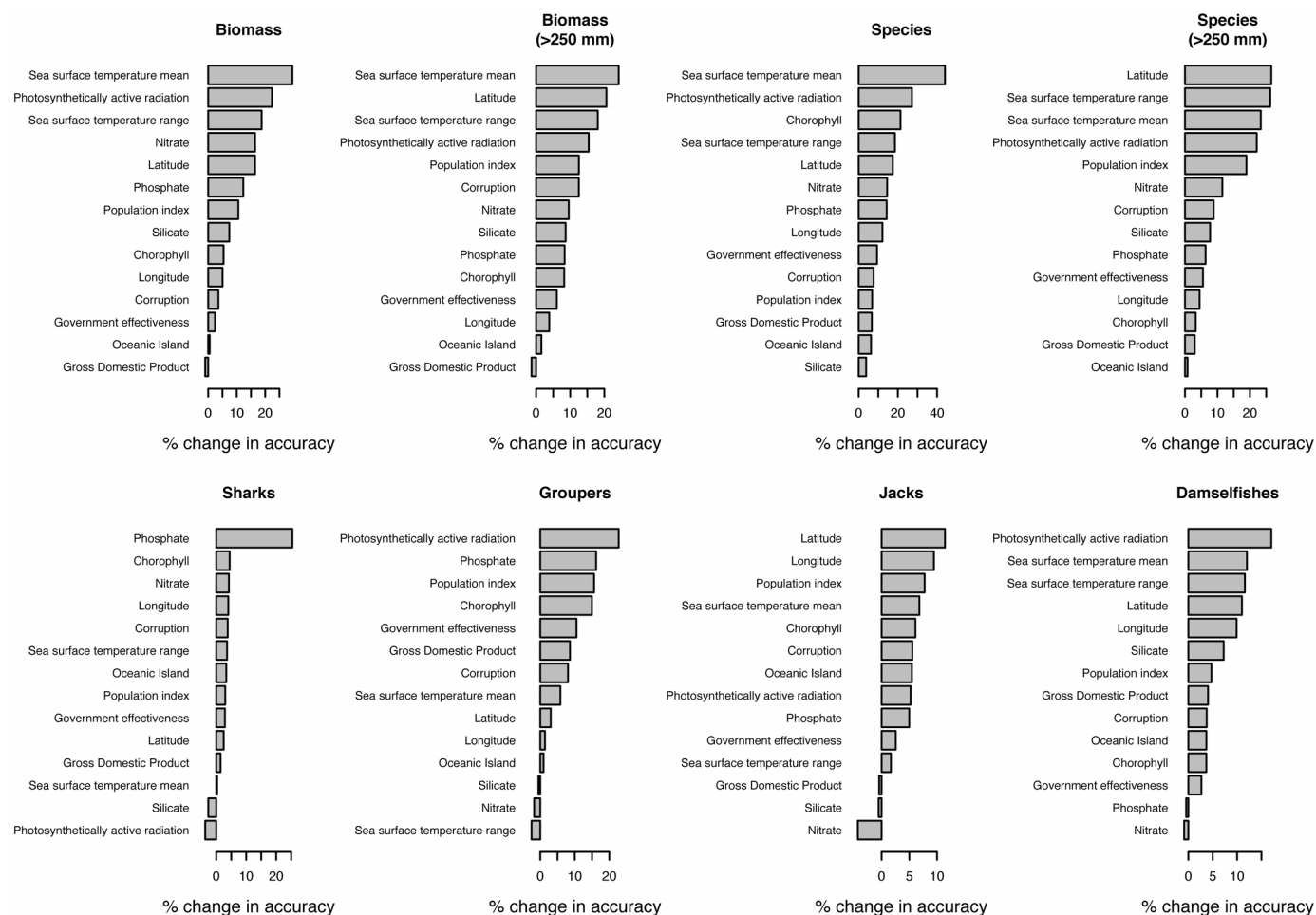
Random forest models were also used to predict values of each metric for fished sites across 5 arcmin grid cells globally, which were then plotted on maps within a coastal buffer. The calculations underlying random forest models used to generate global maps differed from calculations used to predict MPA values in two ways: (1) they were based on 10 rather than 14 environmental covariates, with government effectiveness, corruption, GDP and oceanic island not considered given their small contribution to models (Extended Data Fig. 2) and difficulty in compilation through the full global prediction space; and (2) data for the four fish groups were $\log[x + \text{minimum value for metric}]$ transformed before analysis and back transformed post hoc to compensate for the many zeroes associated with global mapping predictions.

31. Edgar, G. J., Barrett, N. S. & Morton, A. J. Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *J. Exp. Mar. Biol. Ecol.* **308**, 269–290 (2004).
32. CIESIN & CIAT. *Gridded Population of the World, Version 3* (Columbia Univ., 2005).
33. Silverman, B. W. *Density Estimation for Statistics and Data Estimation* (Chapman & Hall, 1986).
34. Kaufmann, D., Kraay, A. & Mastruzzi, M. The Worldwide Governance Indicators: a summary of methodology, data and analytical issues. World Bank Policy Research Working Paper no. 5431, <http://ssrn.com/abstract=1682130> (2010).
35. Tyberghein, L. *et al.* Bio-ORACLE: a global environmental dataset for marine species distribution modeling. *Glob. Ecol. Biogeogr.* **21**, 272–281 (2012).



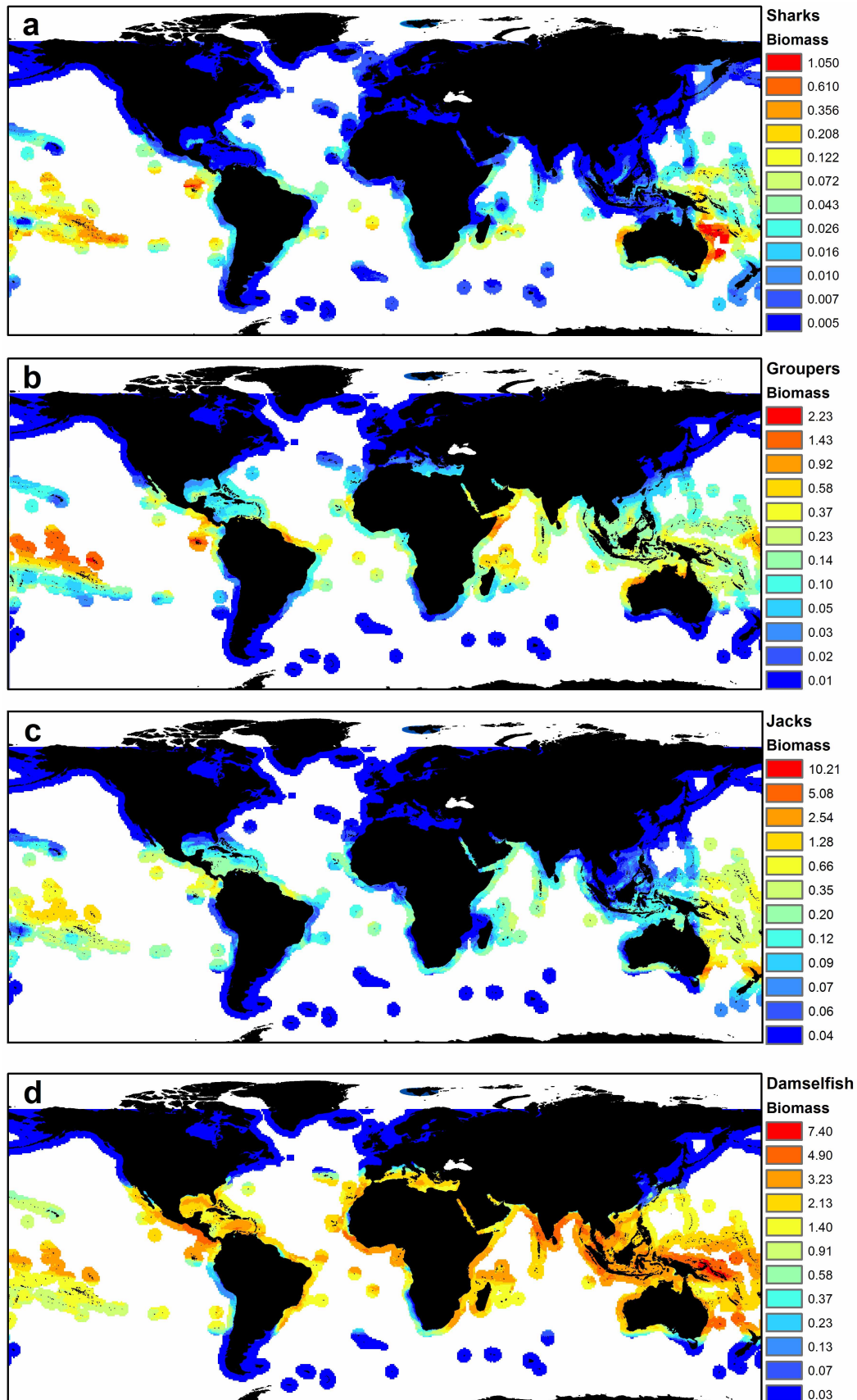
Extended Data Figure 1 | Distribution of sites surveyed. **a**, Number of NEOLI (no take, enforced, old, large and isolated) features at MPAs investigated (coloured circles). MPAs with most NEOLI features are overlaid on top; consequently numerous MPAs with one and two features are not visible. MPAs with five NEOLI features are (1) Cocos, (2) Kermadec Islands,

(3) Malpelo, (4) Middleton Reef; MPAs with four NEOLI features are (5) Elizabeth Reef, (6) Poor Knights Islands, (7) Ship Rock, (8) Tortugas and (9) Tsitsikamma. **b**, All MPA and fished sites surveyed (black circles). Blue shading summarizes the number of sites surveyed within each ecoregion.



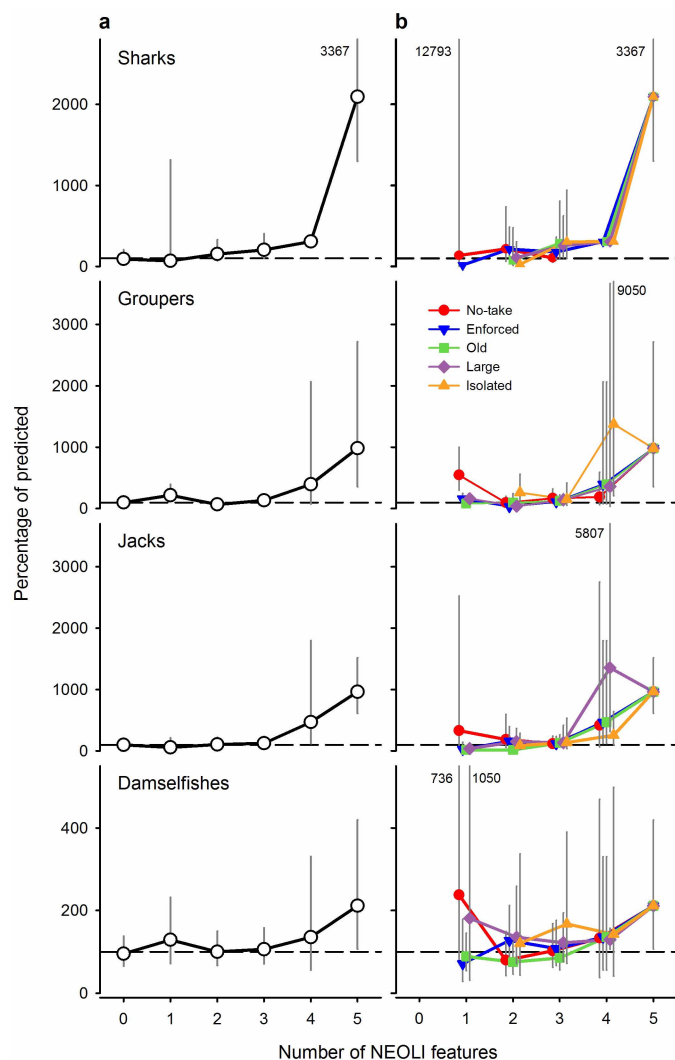
Extended Data Figure 2 | Relative importance of the 14 covariates used in global prediction models developed with random forests. Per cent change in accuracy for a given predictor variable is measured by the change between models that include or do not include that predictor variable, with accuracy

assessed as the mean of the residuals squared. Residuals are based on a cross-validation technique to avoid bias, and the change in accuracy is divided by the standard error for a given tree then averaged across all trees.

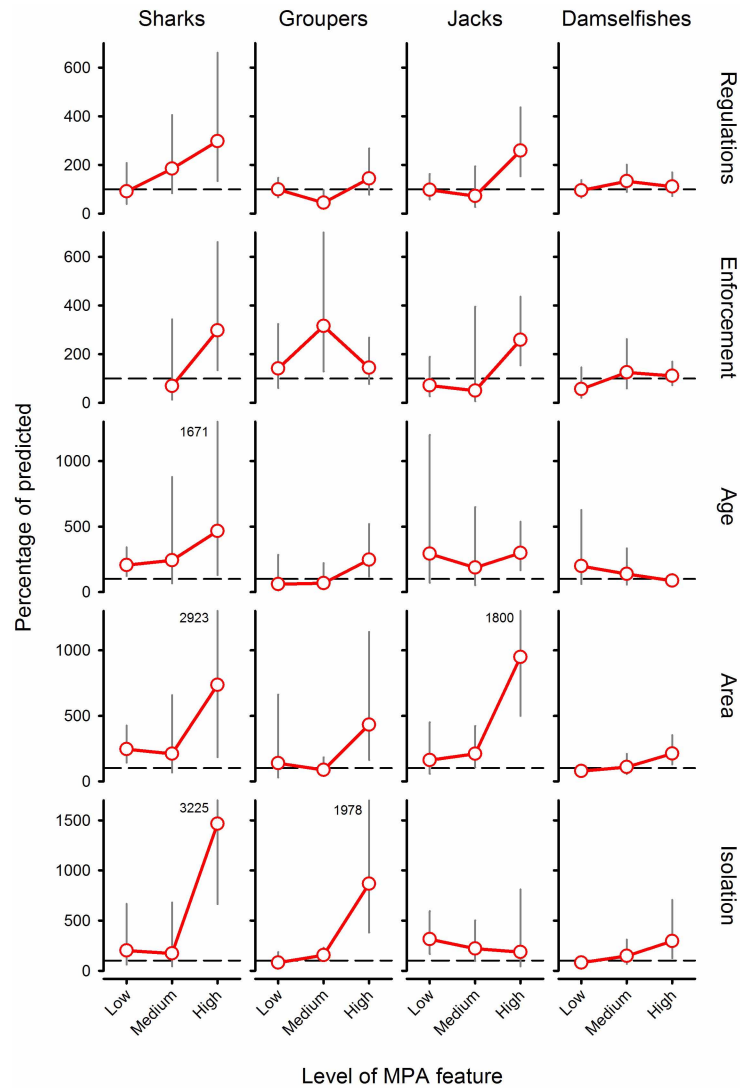


Extended Data Figure 3 | Predicted global distribution of fish biomass (kg per 250 m²) on fished coasts. Predictions are from random forest models developed using data from 1,022 sites in fished locations worldwide. **a**, Sharks.

b, Groupers. **c**, Jacks. **d**, Damselfishes. Note that scales in colour schemes differ among maps, and numbers represent predicted values represented by each colour after smoothing of log-transformed site-level data.



Extended Data Figure 4 | Mean response ratios for MPAs with different number of NEOLI features. Mean ratio values have been back transformed from logs and expressed as percentages with 95% confidence intervals. The number of NEOLI features varies from 0 at sites along fished coastlines to 5 for MPA sites with all NEOLI features. **a**, Plots calculated for sites where sharks, groupers, jacks and damselfishes were present and the subsets of MPAs with different numbers of NEOLI (no take, enforced, old, large, isolated) features. **b**, Mean response ratios for community metrics where each NEOLI feature was included within the set examined. 95% confidence limits that lie off-scale are shown by number. Sample sizes are shown in Extended Data Table 1.



Extended Data Figure 5 | Mean response ratios for the subsets of sites at which sharks, groupers, jacks and damselfishes were observed. Values have been back transformed to per cent, with 100% equivalent to fished coasts, and with 95% confidence intervals. The feature 'regulations' was analysed using data from 82 MPAs that are well enforced; the feature 'enforcement' was analysed using data from 75 MPAs that are no take; and the features 'isolation',

'age' and 'area' were analysed using data from 52 MPAs that are both no take and well enforced. Sharks were not observed in any no-take MPA with low enforcement, so the associated response ratio could not be calculated. 95% confidence limits that lie off-scale are shown by number. Sample sizes are shown in Extended Data Table 1.

Extended Data Table 1 | Sample sizes applied in figures

Figure 2a		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
All metrics	All	76	16	57	39	5	4

Figure 2b		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
All metrics	No-take	7	31	29	4	4	
	Enforced	5	36	32	5	4	
	Old	2	22	31	5	4	
	Large	2	16	16	3	4	
	Isolated		8	9	3	4	

Extended Data Figure 4a		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
Sharks	All	29	3	26	14	1	4
Groupers	All	46	8	36	22	4	4
Jacks	All	50	6	29	27	4	4
Damselfishes	All	62	12	48	36	4	4

Extended Data Figure 4b		No. of NEOLI features					
Metric	MPA subset	0	1	2	3	4	5
Sharks	No-take	2	13	9			4
	Enforced	1	22	10	1		4
	Old		5	8	1		4
	Large		9	9	1		4
	Isolated		3	6	1		4
Groupers	No-take	3	19	14	3		4
	Enforced	2	21	16	4		4
	Old	2	14	18	4		4
	Large	1	11	12	3		4
	Isolated		7	6	2		4
Jacks	No-take	2	12	20	3		4
	Enforced	2	21	21	4		4
	Old	1	10	20	4		4
	Large	1	10	11	2		4
	Isolated		4	9	3		4
Damselfishes	No-take	4	25	26	3		4
	Enforced	4	31	29	4		4
	Old	2	17	28	4		4
	Large	2	14	16	2		4
	Isolated		8	9	3		4

Figure 3		Level of feature		
Metric	Feature	Low	Medium	High
All metrics	Regulations	76	30	52
	Enforcement	9	14	52
	Age	5	19	28
	Area	14	30	8
	Isolation	35	9	8

Extended Data Figure 5		Level of feature		
Metric	Feature	Low	Medium	High
Sharks	Regulations	29	15	23
	Enforcement	0	5	23
	Age	3	12	8
	Area	3	14	6
	Isolation	12	6	5
Groupers	Regulations	46	20	27
	Enforcement	7	9	27
	Age	3	8	16
	Area	6	14	7
	Isolation	16	6	5
Jacks	Regulations	50	19	33
	Enforcement	2	6	33
	Age	3	10	20
	Area	8	19	6
	Isolation	19	6	8
Damselfishes	Regulations	62	28	44
	Enforcement	8	10	44
	Age	3	17	24
	Area	11	26	7
	Isolation	29	7	8

Extended Data Table 2 | Covariates used as predictor variables in global random forest models

Variable	Variable abbreviation	Units	Scale	Reference (if applicable)
Index of population pressure	POP_index	index	2.46 arcmin (4.6 km)	
Government Effectiveness	Govt Eff	index	country	35
Control of Corruption	Corruption	index	country	35
Per capita GDP	GDP	US\$	country	#
Mean nitrate	Bio_nitrate	umol/l	5 arcmin (9.2 km)	36
Mean phosphate	BIO_phosphate	umol/l	5 arcmin (9.2 km)	36
Mean silicate	BIO_silicate	umol/l	5 arcmin (9.2 km)	36
Mean chlorophyll A	BIO_chlomean	mg/m ³	5 arcmin (9.2 km)	36
Photosynthetically active radiation	BIO_parmean	Einstein/m ³ /day	5 arcmin (9.2 km)	36
Mean sea surface temperature	BIO_SST_mean	°C	5 arcmin (9.2 km)	36
Range of sea surface temperature	BIO_SST_range	°C	5 arcmin (9.2 km)	36
Oceanic island isolated from continental shelf	Oceanic island	yes/no		
Site latitude	Latitude	decimal degrees	0.0001°	
Site longitude	Longitude	decimal degrees	0.0001°	

The index of population pressure was calculated by fitting a smoothly tapered surface to each settlement point on a year 2000 world population density grid³² using the quadratic kernel function³³. Populations were screened for a density greater than 1,000 people per 0.04 degree cell, and the search radius was set at 3.959 degrees. This table contains refs 34 and 35.

Per capita GDP was obtained from IMF for 2012 at http://en.wikipedia.org/wiki/List_of_countries_by_GDP_%28PPP%29_per_capita.

Amazon forests maintain consistent canopy structure and greenness during the dry season

Douglas C. Morton¹, Jyoteshwar Nagol^{2,3}, Claudia C. Carabaja^{1,4}, Jacqueline Rosette^{1,2,5}, Michael Palace⁶, Bruce D. Cook¹, Eric F. Vermote¹, David J. Harding¹ & Peter R. J. North⁵

The seasonality of sunlight and rainfall regulates net primary production in tropical forests¹. Previous studies have suggested that light is more limiting than water for tropical forest productivity², consistent with greening of Amazon forests during the dry season in satellite data^{3–7}. We evaluated four potential mechanisms for the seasonal green-up phenomenon, including increases in leaf area^{5–7} or leaf reflectance^{3,4,6}, using a sophisticated radiative transfer model⁸ and independent satellite observations from lidar and optical sensors. Here we show that the apparent green up of Amazon forests in optical remote sensing data resulted from seasonal changes in near-infrared reflectance, an artefact of variations in sun-sensor geometry. Correcting this bidirectional reflectance effect eliminated seasonal changes in surface reflectance, consistent with independent lidar observations and model simulations with unchanging canopy properties. The stability of Amazon forest structure and reflectance over seasonal timescales challenges the paradigm of light-limited net primary production in Amazon forests and enhanced forest growth during drought conditions. Correcting optical remote sensing data for artefacts of sun-sensor geometry is essential to isolate the response of global vegetation to seasonal and interannual climate variability.

The role of Amazon forests in the global carbon budget remains uncertain^{9–12}. Efforts to better constrain net carbon emissions from Amazonia have focused on the magnitude of deforestation carbon fluxes¹³ and the seasonal and interannual variability of Amazon forest productivity^{1,14–16}. Variability in Amazon forest productivity is potentially larger than deforestation emissions on an annual basis, yet remains poorly constrained by field^{14,16} or atmospheric observations^{11,12}. At the centre of this debate is whether tropical forest productivity is more limited by sunlight or rainfall^{3,4,9,14,16,17}. Resolving this issue is critical to reducing uncertainties in the contemporary carbon balance of tropical forests^{10–12,15,16} and the probable response of Amazon forests to climate change^{9,18,19}.

Satellite instruments offer a global perspective on seasonal and interannual changes in vegetation productivity. Indices of fractional vegetation cover derived from satellite data suggest a ‘green up’ over Amazon forests during the dry season^{3,17}, consistent with the light limitation hypothesis². Previous studies have attributed this seasonal greening of Amazon forests to synchronous canopy leaf turnover^{3,4,6}, as young leaves reflect more near-infrared (NIR) light than the older leaves they replace²⁰, or seasonal increases in green leaf area^{5–7}. These leaf-level responses would increase the photosynthetic capacity of Amazon forests, resulting in higher net primary production (NPP) as photosynthetically active radiation (PAR) increases during dry season months^{1,3,4}. Light-limited forests could therefore respond to drought events or other cloud-free periods with an increase in NPP¹⁷, given sufficient access to deep soil moisture²¹.

NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) is the primary data source for studies of the green-up phenomenon,

including the enhanced vegetation index (EVI)^{3,4,7} and leaf area index (LAI) products^{5,7}. The MODIS EVI and LAI products are very sensitive to changes in NIR reflectance^{7,22}. Several mechanisms could generate an increase in the NIR reflectance of Amazon forests (Table 1); increases in MODIS EVI or LAI alone are therefore insufficient to isolate the biophysical basis for the Amazon green up^{6,7}.

We combined three lines of evidence to test the hypotheses that changes in leaf area or leaf reflectance drive the appearance of a seasonal green up of forests in southern Amazonia. First, we synthesized a three-dimensional Amazon forest (Extended Data Fig. 1) and used the FLIGHT radiative transfer model⁸ to simulate changes in lidar and optical remote sensing metrics from seasonal variability in leaf and litter reflectances, leaf area, and sun-sensor geometry during the June to October dry season (Table 1). Second, we analysed lidar waveform centroid relative height (WCRH)²³ and 1064 nm retroreflectance measurements from the Geosciences Laser Altimeter System (GLAS) lidar instrument onboard NASA’s Ice, Cloud, and land Elevation satellite (ICESat)²⁴. GLAS footprint pairs were selected from adjacent orbits in June and October across the southern Amazon. Finally, we reprocessed daily surface reflectance data at 1 km spatial resolution from the Terra and Aqua MODIS sensors to eliminate artefacts from changes in sun-sensor geometry. The combination of theoretical modelling and independent lidar and optical satellite observations in this study generated multiple lines of evidence to better characterize the apparent green-up phenomenon. We found no evidence for consistent seasonal changes in canopy structure or reflectance properties of southern Amazon forests from lidar and optical remote sensing observations. Our approach to constrain seasonal variability in NIR reflectance has wide-ranging implications

Table 1 | Lidar and optical remote sensing metrics are sensitive to possible leaf-level mechanisms for a seasonal green up of Amazon forests

Model simulations	WCRH	EVI	NDVI
June*	0.52	0.40	0.87
October†			
Sun-sensor geometry		0.43	0.81
ρ_{Litter} (+20–25%)‡	0.48	0.48	0.83
ρ_{Leaf} (+10%)	0.52	0.54	0.85
LAI (+2)	0.58	0.43	0.83
LAI + ρ_{Litter}	0.56	0.46	0.84
LAI + ρ_{Leaf}	0.58	0.55	0.87
LAI + ρ_{Leaf} + ρ_{Litter}	0.56	0.59	0.88

Model simulations evaluated the impact of seasonal changes in sun-sensor geometry and increases in litter reflectance (ρ_{Litter}), leaf reflectance (ρ_{Leaf}), and leaf area index (LAI) on GLAS lidar WCRH and MODIS vegetation indices (EVI, NDVI).

* Amazon forest conditions in June for FLIGHT radiative transfer model simulations of GLAS lidar (LAI = 4.5, 1064 nm ρ_{Leaf} = 0.39, and 1064 nm ρ_{Litter} = 0.34) and MODIS surface reflectance, uncorrected for sun-sensor geometry (LAI = 4.5, 860 nm ρ_{Leaf} = 0.39, 860 nm ρ_{Litter} = 0.30; and mean Terra MODIS solar zenith angle (θ_s) = 40°, view zenith angle (θ_v) = 15°, and relative azimuth (ϕ) = 60°).

† October viewing conditions for MODIS simulations were θ_s = 20°, θ_v = 15°, ϕ = 0°.

‡ ρ_{Litter} was increased by 20% for lidar (0.54) and 25% (0.55) for optical simulations with FLIGHT (see Extended Data Fig. 1).

¹NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ²University of Maryland, College Park, Department of Geographical Sciences, College Park, Maryland 20742, USA. ³Global Land Cover Facility, College Park, Maryland 20740, USA. ⁴Sigma Space Corporation, Lanham, Maryland 20706, USA. ⁵Swansea University, Department of Geography, Singleton Park, Swansea SA2 8PP, UK. ⁶Earth System Research Center, University of New Hampshire, Durham, New Hampshire 03824, USA.

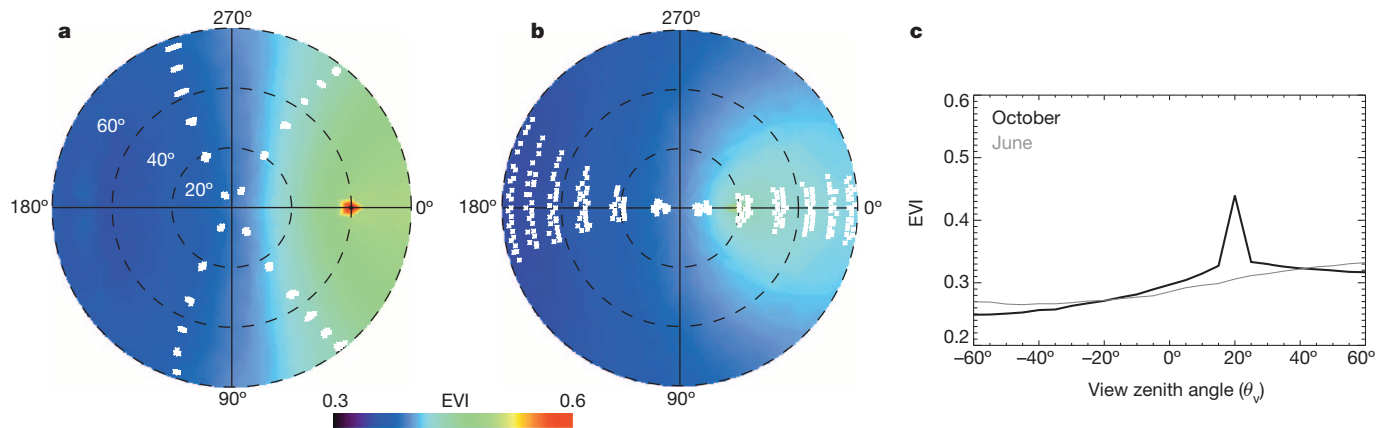


Figure 1 | Seasonal changes in sun-sensor geometry generate the appearance of a green up in MODIS observations of Amazon forest between June and October. **a–c**, Modelled directional anisotropy of the MODIS enhanced vegetation index (EVI) is stronger in June (**a**) than October (**b**), but the realized bidirectional reflectance effect is greater in October (**b**, **c**) when the

MODIS instruments sample in the principal plane ($\phi = 0^\circ$, $\phi = 180^\circ$). White dots in **a** and **b** indicate the relative azimuth angle of Terra and Aqua MODIS observations. The peak in modelled October EVI near $\theta_v = 20^\circ$ indicates the hot-spot effect from the alignment of solar illumination and sensor observation angles.

for *in situ*, airborne and satellite-based studies of vegetation phenology and canopy structure that rely on NIR spectral information.

Three potential biophysical mechanisms for a green up of Amazon forests generated unique responses in modelled lidar and optical remote sensing metrics (Table 1 and Extended Data Table 1). Increasing litter reflectance, consistent with greater litter depth or lower litter moisture during the dry season, decreased the WCRH and increased EVI. Increasing leaf reflectance had negligible effects on WCRH based on compensating increases in both single and multiple scattering of NIR photons, but led to a 32–35% increase in EVI relative to simulations with unchanging canopy properties. Adding LAI increased both WCRH and EVI, as did combinations of increasing LAI, leaf reflectance and litter reflectance. A fourth mechanism, changes in sun-sensor geometry, only influenced simulations of MODIS reflectance data. During the Amazon dry season, the relative azimuth angle (ϕ) of MODIS observations decreases until the sensor approaches the principal plane ($\phi = 0^\circ$, $\phi = 180^\circ$) following the September equinox (Fig. 1 and Extended Data Fig. 3). Changes in sun-sensor geometry alone increased EVI across the range of view zenith angles selected for most land applications (Fig. 1). Modelled changes were greatest near the hot spot, although the magnitude of the hot-spot effect is difficult to reproduce in a theoretical model²⁵. Seasonal changes in sun-sensor geometry decreased the modelled normalized difference vegetation index (NDVI) due to an increase in red reflectance in the hot spot (Table 1 and Extended Data Fig. 4). The magnitude of changes in modelled and observed MODIS NDVI was smaller than for EVI (see Supplementary Information). Results of the FLIGHT model simulations indicated that combined information from lidar

and optical satellite sensors could isolate leaf or litter mechanisms for seasonal greening from artefacts of changing sun-sensor geometry in MODIS observations.

The distributions of relative heights and apparent reflectance estimates from GLAS lidar observations did not increase between June and October across the southern Amazon (Fig. 2). Consistent WCRH values at the beginning and end of the dry season could occur despite increasing leaf reflectance (Table 1). However, the distribution of the most reflective lidar returns was also similar between June and October, indicating that the overall reflectivity of the forests at 1064 nm did not increase. Combined, height and reflectance estimates from the GLAS lidar instrument do not provide consistent regional evidence for a leaf-level mechanism for seasonal greening of Amazon forests.

The spatial patterns of lidar metrics also differed from previous maps of the Amazon green-up phenomenon. Distributions of GLAS lidar WCRH in June and October were similar across the southern Amazon (Extended Data Fig. 5). As a relative metric, WCRH is more robust to residual atmospheric contamination from biomass burning that may influence lidar apparent reflectance estimates over Amazon forests. The spatial pattern of WCRH estimates during the dry season highlights greater penetration of lidar energy in forests with lower mean annual precipitation (Extended Data Fig. 6). We speculate that lower WCRH estimates in these drier forest regions may arise from greater vertical heterogeneity from emergent trees or higher gap fraction because of moisture competition. A higher degree of within-canopy shadowing in forests with low WCRH could exacerbate the directional anisotropy in surface reflectance from optical sensors such as MODIS.

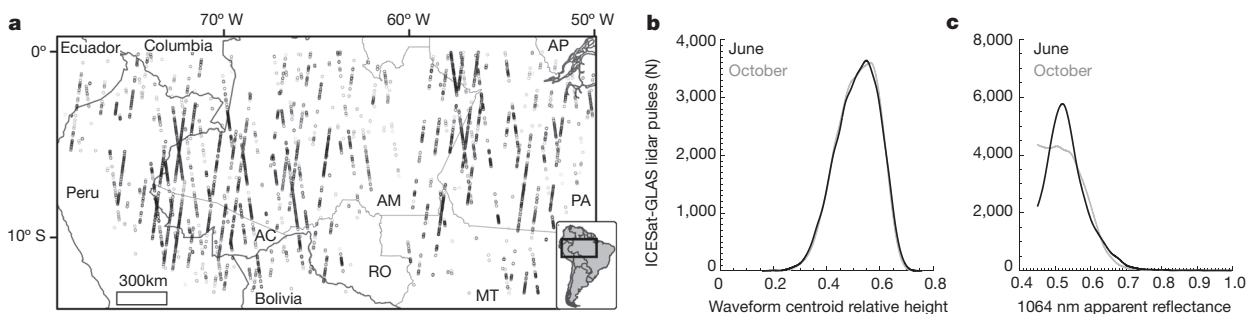


Figure 2 | Independent GLAS lidar observations indicate consistent canopy properties between June and October across southern Amazonia. **a**, Map of GLAS lidar footprint pairs from adjacent orbits during June (L3c and L3f) and October (L3a and L3i) laser operation periods ($n = 74,638$). **b**, **c**, Distributions of lidar waveform centroid relative heights (WCRH) (**b**) and maximum

apparent reflectance (**c**) (1064-nm retrorreflectance) did not increase between early and late dry season laser operation periods. States within the Brazilian Amazon are abbreviated in **a**. Acre, AC; Amazonas, AM; Amapá, AP; Mato Grosso, MT; Pará, PA; Rondônia, RO.

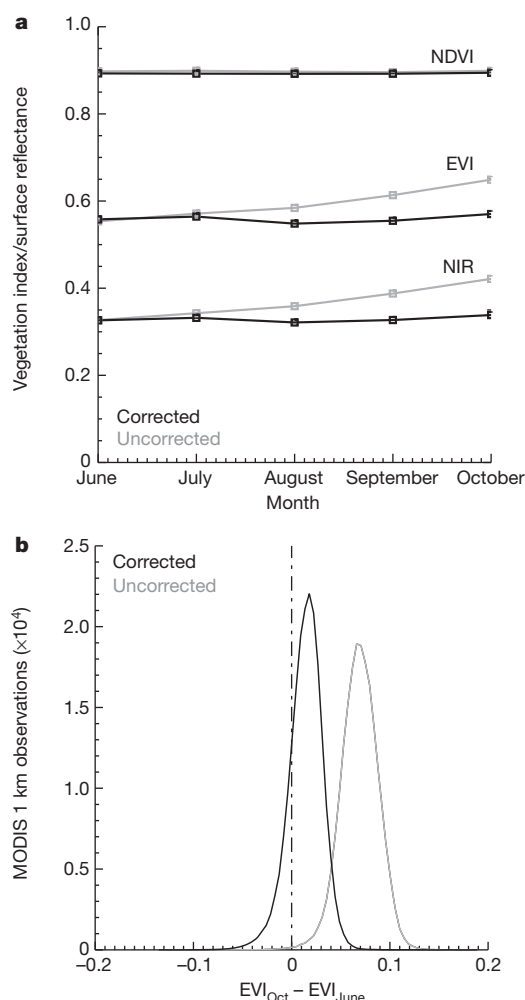


Figure 3 | Seasonal increases in MODIS near-infrared reflectance (NIR) and enhanced vegetation index (EVI) data were eliminated after normalizing sun-sensor geometry during the June to October dry season. **a**, Seasonal profile of the normalized difference vegetation index (NDVI), EVI and NIR for uncorrected (grey) and corrected (black) MODIS data in southern Amazonia. Values indicate the upper quartile median monthly NDVI, EVI and NIR of all Terra and Aqua MODIS observations in 2003–08 ($n = 197,651$). **b**, Per-pixel changes in uncorrected (grey) and corrected (black) MODIS EVI between October and June for forested areas in southern Amazonia.

Removing artefacts of changing sun-sensor geometry in MODIS red and NIR reflectance data eliminated the appearance of a green up in MODIS EVI during dry season months (Fig. 3 and Extended Data Fig. 7). Previous green-up studies with MODIS EVI or LAI data did not explicitly account for changing viewing and illumination conditions^{3,7,17} or did not normalize to a constant sun-sensor geometry for all observations⁴. Corrected EVI data also eliminated the appearance of interannual variability during 2003–08 (Extended Data Fig. 8). Artefacts from changes in sun-sensor geometry during the dry season may therefore explain previous reports of higher EVI values over Amazon forests during drought conditions¹⁷. Indeed, lower cloud cover during drought years increases the probability of observing forests in the southern Amazon when MODIS observations are closer to the principal plane (Extended Data Fig. 9).

Seasonal increases in NIR reflectance drive the apparent green up of Amazon forests, based on a reduction in self-shading near the September equinox that is amplified in MODIS EVI data by changes in sun-sensor geometry. Seasonal changes in NIR reflectance and EVI do not correspond directly to changes in leaf reflectance, leaf area or vegetation productivity. However, the sensitivity of uncorrected MODIS EVI to seasonal illumination may serve as a useful proxy for variability in the

fraction of absorbed PAR (FAPAR). This relationship between EVI and FAPAR may explain the positive correlation between MODIS EVI and net ecosystem exchange (NEE) in Amazon forests³ and other biomes²⁶. The potential response of Amazon forests to increasing PAR in the dry season, as measured at the top of canopy¹, is therefore moderated by seasonal variability in FAPAR from the reduction in self-shading between the June solstice and September equinox.

Satellite-based evidence for consistent reflectance of Amazon forest canopies over seasonal and annual time scales reframes the debate over climate controls on Amazon forest productivity. Leaf turnover may still be important for increased productivity of individual trees during dry season months^{4,6,19,20}, but GLAS and MODIS observations do not support the hypotheses that changes in canopy reflectance properties from leaf phenology are consistent at larger spatial scales (≥ 1 km) across the southern Amazon. Direct observations of this fine-scale variability in Amazon forest NPP may yet be possible from satellite-based estimates of vegetation fluorescence²⁷ or light-use efficiency²⁸, provided that the fraction of sunlit and shaded leaf area can be retrieved or modelled simultaneously. Our results support the initial interpretations of eddy covariance studies¹, rainfall exclusion experiments²⁹ and plot measurements¹⁴ in Amazon forests that seasonal moisture availability governs the balance between photosynthesis and respiration in Amazon forests.

METHODS SUMMARY

A 1 km² Amazon forest was synthesized using field measurements of forest structure and reflectance properties (see Extended Data Fig. 1). Tree locations and attributes were used to generate a three-dimensional forest scene in the FLIGHT radiative transfer model⁸. FLIGHT model simulations quantified the impact of individual and combined seasonal changes in leaf reflectance, litter reflectance, LAI, and MODIS relative azimuth angle on optical and lidar remote sensing metrics (see Extended Data Figs 1 and 2 and Extended Data Table 1). Simulated seasonal changes in remote sensing metrics were compared with GLAS and MODIS observations.

ICESat-GLAS lidar data over southern Amazon forests were selected from Laser 3 operation periods in the early (L3c, May to June 2005; L3f, May to June 2006) and late dry seasons (L3a, October 2004; L3i, October 2007). GLAS waveform characteristics were derived from the GLA09 and GLA14 standard products (data access: <http://nsidc.org>). High-quality data from adjacent orbits in June and October were used to assess seasonal changes in lidar apparent reflectance and WCRH²³ ($n = 74,638$).

Terra and Aqua MODIS daily 1 km surface reflectance data from 2003–08 were acquired for two $10^\circ \times 10^\circ$ spatial tiles (H11V09 and H12V09, Collection 5 MOD09GA and MYD09GA, data access: <https://lpdaac.usgs.gov/>). Each month, high-quality Terra and Aqua MODIS data were combined to estimate the bidirectional reflectance distribution function (BRDF) and perform a per-pixel correction for variations in sun-sensor geometry²⁵. A threshold of ≥ 20 observations per pixel, per month was set to ensure a robust statistical inversion ($n = 197,651$). BRDF parameters were used to normalize red and NIR reflectance values to constant solar ($\theta_s = 30^\circ$) and view geometries ($\theta_v = 0^\circ$). NDVI and a two-band version of EVI³⁰ were calculated using the normalized red and NIR reflectance data. Uncertainties in normalized MODIS data were assessed using a Monte Carlo sensitivity analysis.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 December 2012; accepted 8 January 2014.

Published online 5 February 2014.

1. Saleska, S. R. *et al.* Carbon in Amazon forests: unexpected seasonal fluxes and disturbance-induced losses. *Science* **302**, 1554–1557 (2003).
2. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
3. Huete, A. R. *et al.* Amazon rainforests green-up with sunlight in dry season. *Geophys. Res. Lett.* **33**, L06405 (2006).
4. Brando, P. M. *et al.* Seasonal and interannual variability of climate and vegetation indices across the Amazon. *Proc. Natl Acad. Sci. USA* **107**, 14685–14690 (2010).
5. Myneni, R. B. *et al.* Large seasonal swings in leaf area of Amazon rainforests. *Proc. Natl Acad. Sci. USA* **104**, 4820–4823 (2007).

6. Doughty, C. E. & Goulden, M. L. Seasonal patterns of tropical forest leaf area index and CO₂ exchange. *J. Geophys. Res.* **113**, G00B06 (2008).
7. Samanta, A. *et al.* Seasonal changes in leaf area of Amazon forests from leaf flushing and abscission. *J. Geophys. Res.* **117**, G01015 (2012).
8. North, P. R. J., Rosette, J. A. B., Suarez, J. C. & Los, S. O. A Monte Carlo radiative transfer model of satellite waveform LiDAR. *Int. J. Remote Sens.* **31**, 1343–1358 (2010).
9. Davidson, E. A. *et al.* The Amazon basin in transition. *Nature* **481**, 321–328 (2012).
10. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
11. Chevallier, F. *et al.* Global CO₂ fluxes inferred from surface air-sample measurements and from TCCON retrievals of the CO₂ total column. *Geophys. Res. Lett.* **38**, L24810 (2011).
12. Gatti, L. V. *et al.* Vertical profiles of CO₂ above eastern Amazonia suggest a net carbon flux to the atmosphere and balanced biosphere between 2000 and 2009. *Tellus B Chem. Phys. Meteorol.* **62**, 581–594 (2010).
13. van der Werf, G. R. *et al.* CO₂ emissions from deforestation. *Nature Geosci.* **2**, 737–738 (2009).
14. Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
15. Le Quéré, C. *et al.* Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2**, 831–836 (2009).
16. Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F. & Nepstad, D. The 2010 Amazon drought. *Science* **331**, 554 (2011).
17. Saleska, S. R., Didan, K., Huete, A. R. & da Rocha, H. R. Amazon forests green-up during 2005 drought. *Science* **318**, 612–612 (2007).
18. Zelazowski, P., Malhi, Y., Huntingford, C., Sitch, S. & Fisher, J. B. Changes in the potential distribution of humid tropical forests on a warmer planet. *Phil. Trans. Royal Soc. A* **369**, 137–160 (2011).
19. Kim, Y. *et al.* Seasonal carbon dynamics and water fluxes in an Amazon rainforest. *Glob. Change Biol.* **18**, 1322–1334 (2012).
20. Toomey, M., Roberts, D. A. & Nelson, B. The influence of epiphylls on remote sensing of humid forests. *Remote Sens. Environ.* **113**, 1787–1798 (2009).
21. Nepstad, D. C. *et al.* The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. *Nature* **372**, 666–669 (1994).
22. Galvão, L. S. *et al.* On intra-annual EVI variability in the dry season of tropical forest: a case study with MODIS and hyperspectral data. *Remote Sens. Environ.* **115**, 2350–2359 (2011).
23. Carabajal, C. C. & Harding, D. J. SRTM C-Band and ICESat laser altimetry elevation comparisons as a function of tree cover and relief. *Photogramm. Eng. Remote Sensing* **72**, 287–298 (2006).
24. Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D. & DiMarzio, J. P. Overview of the ICESat mission. *Geophys. Res. Lett.* **32**, L21S01 (2005).
25. Bréon, F.-M., Maignan, F., Leroy, M. & Grant, I. Analysis of hot spot directional signatures measured from space. *J. Geophys. Res.* **107**, <http://dx.doi.org/10.1029/2001JD001094> (2002).
26. Rahman, A. F., Sims, D. A., Cordova, V. D. & El-Masri, B. Z. Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes. *Geophys. Res. Lett.* **32**, L19404 (2005).
27. Joiner, J., Yoshida, Y., Vasilkov, A. P., Corp, L. A. & Middleton, E. M. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences* **8**, 637–651 (2011).
28. Hilker, T. *et al.* Inferring terrestrial photosynthetic light use efficiency of temperate ecosystems from space. *J. Geophys. Res.* **116**, G03014 (2011).
29. Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. & Cardinot, G. Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology* **88**, 2259–2269 (2007).
30. Jiang, Z., Huete, A. R., Didan, K. & Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **112**, 3833–3845 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements Funding for this research was provided by NASA, the NASA Postdoctoral Program and the NERC National Centre for Earth Observation. We thank G. P. Asner for providing Amazon leaf reflectance data.

Author Contributions D.C.M. and J.N. designed the experiment. D.C.M., J.N. and E.F.V. analysed MODIS data. D.C.M., C.C.C., B.D.C. and D.J.H. analysed ICESat data. M.P. developed the synthetic forest model. D.C.M., J.N., J.R., B.D.C. and P.R.J.N. contributed to radiative transfer model simulations. D.C.M. wrote the manuscript, and all authors contributed material to the final version.

Author Information Data from NASA's ICESat, Terra, and Aqua satellites are archived at <http://www.nsidc.org> and <http://lpdaac.usgs.gov>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.M. (douglas.morton@nasa.gov).

METHODS

Synthetic Amazon forest. We synthesized an Amazon forest scene for radiative transfer model simulations using a three-dimensional canopy model and field measurements of forest structure (Extended Data Fig. 1). The modelled forest was generated using allometric equations for the relationships between diameter at breast height (*dbh*), canopy height (*h*), crown width (*C_w*), and crown depth (*C_d*). For trees >20-cm diameter at breast height, we used published allometric equations for the eastern Amazon^{31–33}. Log-linear relationships derived from field measurements of large trees provided unrealistic values for smaller trees, especially stems <10-cm diameter at breast height. Therefore, we adjusted the allometric relationships to better represent smaller diameter stems in the synthetic forest scene.

In the forest understory, crown dimensions are more strongly correlated with *h* than *dbh*^{34,35}. We adjusted the original *h*-*dbh* relationships for small trees, using the large-tree allometry estimates for trees 10–25-cm diameter at breast height. These original estimates were fit with a power-law function to estimate the *h* (m) of trees with diameter at breast height < 20 cm (Extended Data Fig. 1):

$$h = 1.37dbh^{0.88}$$

Adjusted *h* estimates for small diameter trees were then used to develop revised estimates of *C_w* and *C_d*. The resulting *h*-*C_w* relationship was close to the reported values for a range of understory species in Panama³⁴:

$$C_w = 0.47h^{0.934}$$

A revised estimate of *C_d* for small trees was derived in similar fashion, based on a power-law fit of the *h*-*C_d* relationship for trees 25–50-m in height. Larger trees were selected for estimating *C_d* in small trees because the *h*-*C_d* relationship for trees 10–25-m was asymptotic:

$$C_d = 0.11h^{1.5}$$

This approach resulted in a flattening of tree crowns in the understory, with *C_d* estimates of 10, 30, and 60 cm for trees 1, 2 and 3 m in height, respectively.

We synthesized an Amazon forest scene with 8% gap fraction and 200 trees per hectare (ha), similar to satellite and field-based studies in eastern Amazonia, respectively^{32,36}. Trees were drawn at random from a Weibull distribution of diameter at breast height sizes to populate a 1 × 1 km forest scene (Extended Data Fig. 1). Weibull distribution parameters ($\beta = 1$, $\alpha = 27.5$ -cm diameter at breast height) were based on an extensive survey (392 ha) of diameter at breast height values in the Tapajos National Forest near Santarém, Pará^{31,36}. Trees were located in *x,y* space using a simple spacing algorithm. If more than half of the crown radius of the new individual overlapped with half of the crown radius of an existing tree, the tree was excluded and a new sample was drawn until the stem density reached 20,000 individuals >1-cm diameter at breast height per km². Conceptually, this approach incorporates the concepts of shading (large trees exclude small trees) and gap dynamics (small trees indicate the absence of large trees). The *x,y* location and properties of each stem (*dbh*, *h*, *C_w*, *C_d*) were used to create a three-dimensional forest scene in the FLIGHT model⁸. We estimated mean total aboveground biomass of 312 megagrams carbon per ha for the synthetic forest scene based on an independent height-diameter allometry³⁷, similar to forest inventory data³⁶ and satellite-based estimates^{38,39} of above-ground biomass in the eastern Amazon.

FLIGHT model simulations. We simulated the impact of seasonal changes in Amazon forest structure and reflectance properties on optical and lidar remote sensing metrics using FLIGHT, a three-dimensional model of light transport based on a Monte Carlo simulation of radiative transfer^{8,40}. FLIGHT uses photon trajectories within a canopy to evaluate solar bidirectional reflectance and lidar backscatter retroreflectance (reflectance at 0° phase angle). The model simulates the chain of scattering events along each photon trajectory, including multiple scattering between canopy elements and the ground surface. Atmospheric transmittance is simulated using fixed conditions for atmospheric trace gases and aerosols. Amazon tree crowns were modelled as elliptical shells, and field measurements⁴¹ of reflectance and transmittance (Extended Data Fig. 1) were assigned to scattering elements based on the proportion of leaves (85%) and branches (15%) within each crown. To simulate seasonal increases in LAI, only the foliar area was increased in each canopy, thereby decreasing the relative fraction of bark and branches.

Optical and lidar radiative transfer simulations quantified the impact of seasonal changes in leaf reflectance (ρ_{Leaf} , $\pm 10\%$), litter reflectance (ρ_{Litter} , ± 10 –25%), and LAI (+1, +2) on remote sensing metrics. Multiple changes in seasonal forest structure and reflectance were simulated using combinations of ρ_{Leaf} , ρ_{Litter} , and LAI changes under October illumination conditions (Table 1 and Extended Data Table 1). Optical simulations also considered MODIS viewing geometry and seasonal variability in solar illumination, based on the location of

the simulated Amazon forest (2.65° S, 54.95° W). FLIGHT simulations with varying relative azimuth angle (ϕ , 0°–180°), view zenith angle (θ_v , 0°–60°), and solar zenith angle (θ_s , 20°–40°) provided a more complete picture of the directional anisotropy of surface reflectance from seasonal illumination conditions. These three-dimensional plots of the bidirectional reflectance effect were used to illustrate the impact of seasonal changes in sun-sensor geometry on optical remote sensing data (Fig. 1 and Extended Data Fig. 4).

Lidar radiative transfer simulations with FLIGHT incorporated GLAS instrument characteristics, including nadir incidence angle, 5-ns pulse duration, 57 mJ pulse energy (Laser 3a), 600 km sensor altitude and 450 μrad detector field of view. In addition, a constant 5° slope was added to the Amazon forest scene to simulate the effect of microtopography on laser energy returned from the ground. The distribution of modelled WCRH for the synthetic Amazon forest was similar to GLAS data (Fig. 2 and Extended Data Fig. 2a). However, lidar simulations were conducted using a single footprint to isolate the influence of individual scene modifications on lidar radiative transfer (Extended Data Fig. 2b).

Random background noise, similar to noise levels in GLAS lidar waveforms, was added to FLIGHT lidar waveforms to simulate a realistic signal returned to the detector. Noise levels were consistent with the potential contribution from solar photons to return waveform energy, based on FLIGHT simulations of the solar flux at 1064 nm for solar noon at the equator under June and October illumination conditions (0.8–2.5%). Solar noise constitutes a constant offset term in the vertical lidar waveform and does not impact the relative metrics derived from modelled or observed lidar waveforms. The extent of the forest canopy was determined by setting an amplitude threshold above the background noise, based on the signal-to-noise characteristics of the waveform. A Gaussian smoothing algorithm was used as a low-pass filter before computing lidar metrics, following established waveform-processing methods⁴². The width of the smoothing Gaussian was chosen as the half-width at full maximum of the outgoing laser pulse to avoid over-smoothing the waveform data. The ground peak maximum was detected by determining the zero-crossing point of the first derivative of the smoothed waveform.

ICESat-GLAS lidar analysis. GLAS lidar data provided an independent estimate of seasonal changes in forest structure and canopy reflectance. The GLAS instrument is a large footprint, full-waveform lidar^{24,43}. Each lidar waveform contains information on the height and vertical distribution of scattering elements within the lidar footprint, and the integrated energy from the return waveform can be normalized by the outgoing laser energy to estimate apparent reflectance at 1064 nm. Apparent reflectance corresponds to the footprint retroreflectance multiplied by the square of atmospheric transmission to account for the two-way travel path. The waveform centroid relative height (WCRH)²³ was used in this study to assess changes in the relative height of median energy within the waveform:

$$WCRH = \frac{(CE - S_e E)}{(S_s E - S_e E)}$$

where *CE* corresponds to the waveform centroid elevation, *S_sE* is the signal start elevation and *S_eE* is the signal end elevation as determined by thresholds defined for the alternate (land) waveform processing schemes. GLAS footprints over Amazon forest cover types were selected based on time series of dry season MODIS NDVI data⁴⁴ to select forests >1 km from non-forest cover types and burned forest areas in 2010 (ref. 45). In addition, topographic data from the Shuttle Radar Topography Mission at 90-m resolution (data access: <http://hydrosheds.cr.usgs.gov>) were used to exclude forested areas in the southern Amazon above 500-m elevation or on slopes >10%.

We selected four Laser 3 operation periods for the GLAS analysis. Two laser operation periods captured early dry season conditions (L3c, May to June 2005; L3f, May to June 2006), and two laser operation periods covered the late dry season across the southern Amazon (L3a, October 2004; L3i, October 2007). Data from L3a, L3c, L3f and L3i also had similar footprint characteristics (diameter, circularity and radial energy distribution). Differences in Laser 3 transmit energies between L3a and L3i do not impact the estimates of WCRH and apparent reflectance because these metrics are normalized by the transmit energy. The GLA09 and GLA14 standard data products (data access: <http://nsidc.org>) were used to assemble information on the location, laser energy, atmospheric backscatter, apparent reflectance and waveform characteristics for GLAS footprints over Amazon forests. GLAS data parameters were used to screen lidar footprints with cloud contamination, weak or saturated energies, or specular reflections from water. Specifically, we used the offset from the digital elevation model (<70 m), pulse width (<100 m), WCRH (<1), apparent reflectance (>0.45, <1), and a threshold value in the integrated backscatter (<3.5 × 10⁻⁶) to select cloud-free, high-quality ICESat data. In addition, we used MODIS aerosol optical thickness (AOT <0.09) data from the MOD09CMG product (E. Vermote, unpublished data) from the same date of GLAS observations to further screen the lidar data for high aerosol conditions. Finally, GLAS data from June and October laser periods were ‘paired’ using a 0.02° search radius to select

footprints from adjacent orbits. This paired subset of high-quality GLAS data ensured a similar spatial distribution of lidar footprints across the southern Amazon in early and late dry season conditions. The total number of GLAS footprints was 74,638, equally divided between June (L3c = 22,528, L3f = 14,791) and October (L3a = 20,786, L3i = 16,533).

MODIS analysis. We used Terra and Aqua MODIS data from 2003–08 to evaluate seasonal changes in canopy reflectance at ICESat footprint locations in eastern Amazonia. Daily red and NIR surface reflectances, sun-sensor geometry, and data quality flags were drawn from the Collection 5 MOD09GA and MYD09GA standard products at 1-km spatial resolution for two $10^\circ \times 10^\circ$ spatial tiles (H11V09 and H12V09, data access: <https://lpdaac.usgs.gov>). MODIS data were screened for high aerosols and cloud cover using the MOD09GA and MYD09GA data quality flags, and low-aerosol conditions (AOT < 0.09) were confirmed using AOT estimates retrieved from the MOD09CMG product (E. Vermote, unpublished data). Fixed thresholds were also used to select MODIS observations with low red reflectance (< 0.2), NIR reflectance typical of vegetated surfaces ($0.1 < \text{NIR} < 0.6$), and low view zenith angles (< 60°). Finally, an iterative outlier removal procedure was used to eliminate remaining observations with cloud or aerosol contamination. Outliers were identified as NDVI values > 2 standard deviations from the median NDVI value per pixel, per month. These outlier values were removed, and the process was repeated two more times to remove any residual cloud or aerosol-contaminated data. The resulting data set included 197,651 pixels across the southern Amazon. Larger data volumes could be achieved with alternative approaches for cloud and aerosol screening⁴⁶.

Each month, high-quality MODIS data from Terra and Aqua were combined to estimate the bidirectional reflectance distribution function (BRDF) and perform a per-pixel correction for variations in sun-sensor geometry. The BRDF correction uses Ross-Li-Maignan model, revised to account for the hot-spot effect^{25,47}:

$$\rho(\theta_s, \theta_v, \phi) = k_0 + k_1 F_1(\theta_s, \theta_v, \phi) + k_2 F_2(\theta_s, \theta_v, \phi)$$

Surface reflectance (ρ) at a specified θ_s , θ_v , and ϕ is modelled as a function of volumetric (F_1) and geometric scattering (F_2). The volumetric scattering kernel (F_1) is based on the Ross thick function, corrected for the hot spot, and the geometric kernel (F_2) is based on the Li-sparse reciprocal function²⁵. The model equation can be rewritten to highlight the contribution of the volumetric kernel (k_1/k_0) and geometric kernel (k_2/k_0) to the overall signal:

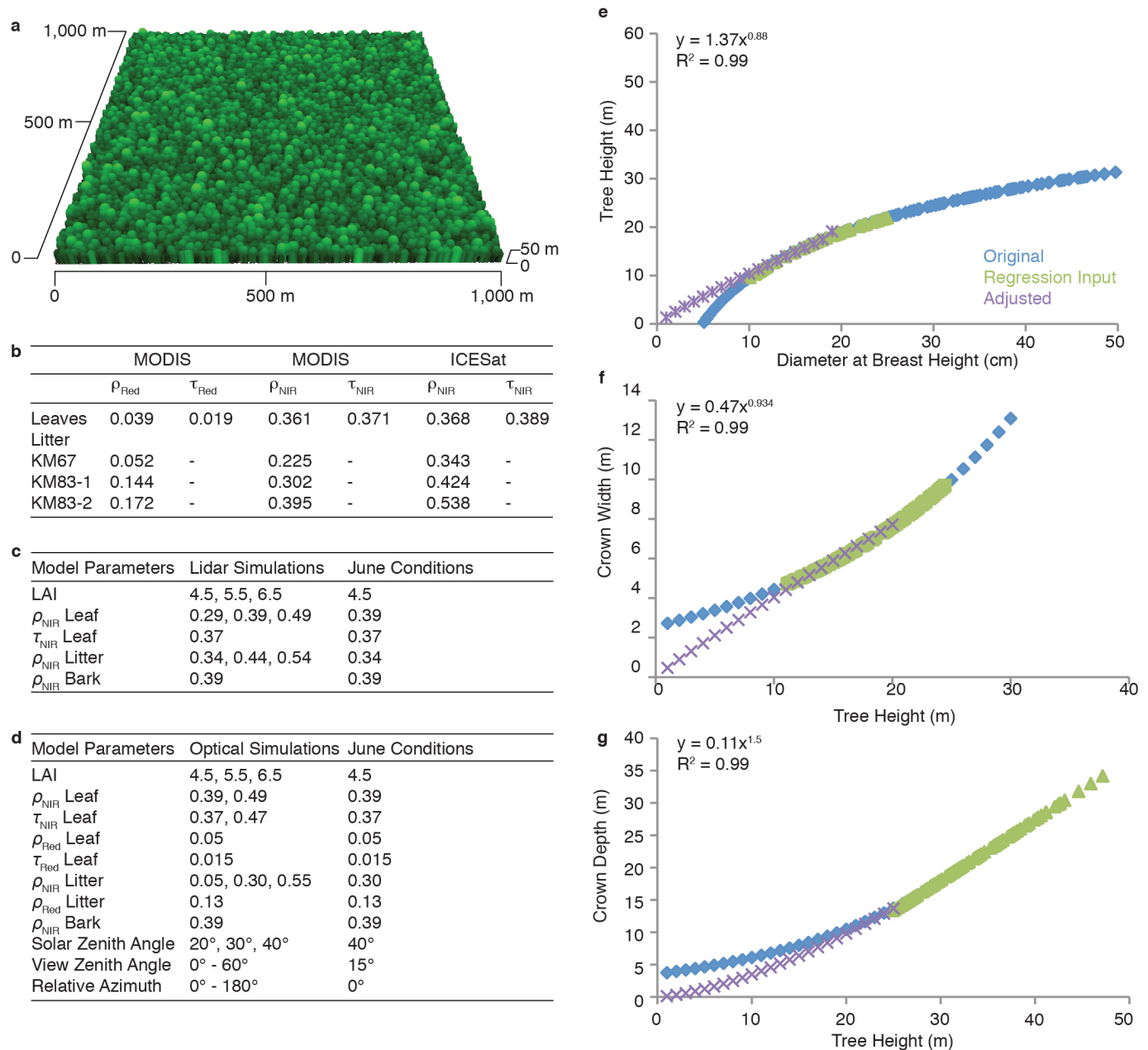
$$= k_0 \left[1 + \frac{k_1}{k_0} F_1(0_s, 0_v, \phi) + \frac{k_2}{k_0} F_2(0_s, 0_v, \phi) \right]$$

The k_0 , k_1 and k_2 parameters were inverted for red and NIR reflectances from all cloud free and low aerosol MODIS data on a per-pixel basis for all dry season months during 2003–08. A threshold of ≥ 20 observations per pixel, per month was set to ensure a robust statistical inversion. These BRDF parameters were then used to normalize both red and NIR reflectance values to a constant solar ($\theta_s = 30^\circ$) and view geometry ($\theta_v = 0^\circ$) for observations in all months. Normalization of sun-sensor geometry in this approach differs from previous efforts to correct MODIS data for BRDF⁴⁸. Finally, NDVI and a two-band version of EVI³⁰ were calculated using the normalized red and NIR reflectance data.

Uncertainty estimation. We estimated the root mean square error (RMSE) of the MODIS surface reflectance product based on mean values of red (0.02) and NIR

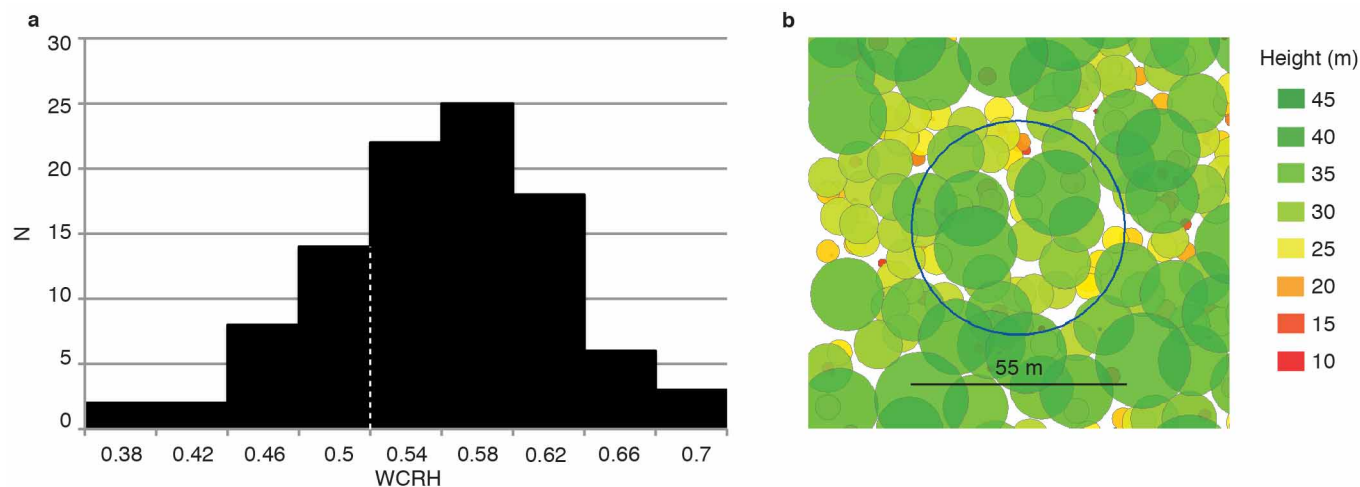
(0.33) reflectance over Amazon forests⁴⁹. Measurement error was propagated to EVI and NDVI using a Monte Carlo sensitivity analysis approach. The RMSE of MODIS EVI (0.56) was 0.036, and the RMSE of MODIS NDVI (0.89) was 0.033. The root mean square model error was approximately one-third as large as measurement error. Assuming independence, total error is approximately 0.035 to 0.037 for both EVI and NDVI, or approximately 6% of mean EVI and 4% of NDVI. Seasonal differences in uncorrected MODIS EVI data exceed this error range (~ 0.1), while seasonal differences in corrected MODIS EVI data do not (see Fig. 3).

31. Palace, M., Keller, M., Asner, G. P., Hagen, S. & Braswell, B. Amazon forest structure from IKONOS satellite data and the automated characterization of forest canopy properties. *Biotropica* **40**, 141–150 (2008).
32. Asner, G. P. *et al.* Estimating canopy structure in an Amazon forest from Laser Range Finder and IKONOS satellite observations. *Biotropica* **34**, 483–492 (2002).
33. Hunter, M. O., Keller, M., Vitoria, D. & Morton, D. C. Tree height and tropical forest biomass estimation. *Biogeosciences Discuss.* **10**, 10491–10529 (2013).
34. King, D. A. Allometry of saplings and understory trees of a Panamanian forest. *Funct. Ecol.* **4**, 27–32 (1990).
35. Osunkoya, O. O. *et al.* Comparative height crown allometry and mechanical design in 22 tree species of Kuala Belalong rainforest, Brunei, Borneo. *Am. J. Bot.* **94**, 1951–1962 (2007).
36. Keller, M., Palace, M. & Hurr, G. C. Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. *For. Ecol. Manage.* **154**, 371–382 (2001).
37. Feldpausch, T. R. *et al.* Height-diameter allometry of tropical forest trees. *Biogeosciences* **8**, 1081–1106 (2011).
38. Saatchi, S. S. *et al.* Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl Acad. Sci. USA* **108**, 9899–9904 (2011).
39. Baccini, A. *et al.* Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Clim. Change* **2**, 182–185 (2012).
40. North, P. R. J. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Trans. Geosci. Rem. Sens.* **34**, 946–956 (1996).
41. Huete, A. R., Ferreira, L. & Miura, T. LBA-ECO LC-19 Field Measurements 2002: Biophysical & Soil Parameters. <http://www.lbaeco.org> (2008).
42. Hofton, M. A., Minster, J. B. & Blair, J. B. Decomposition of laser altimeter waveforms. *IEEE Trans. Geosci. Rem. Sens.* **38**, 1989–1996 (2000).
43. Harding, D. J. & Carabajal, C. C. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophys. Res. Lett.* **32**, L21S10 (2005).
44. Morton, D. C. *et al.* Mapping canopy damage from understory fires in Amazon forests using annual time series of Landsat and MODIS data. *Remote Sens. Environ.* **115**, 1706–1720 (2011).
45. Morton, D. C., Le Page, Y., DeFries, R., Collatz, G. J. & Hurr, G. C. Understorey fire frequency and the fate of burned forests in southern Amazonia. *Phil. Trans. R. Soc. Lond. B* **368** (2013).
46. Hilker, T. *et al.* Remote sensing of tropical ecosystems: atmospheric correction and cloud masking matter. *Remote Sens. Environ.* **127**, 370–384 (2012).
47. Vermote, E. F., Justice, C. & Breon, F.-M. Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectance. *IEEE Trans. Geosci. Rem. Sens.* **47**, 898–908 (2009).
48. Schaaf, C. B. *et al.* First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.* **83**, 135–148 (2002).
49. Vermote, E. F. & Kotchenova, S. Atmospheric correction for the monitoring of land surfaces. *J. Geophys. Res. D Atmospheres* **113**, D23S90 (2008).



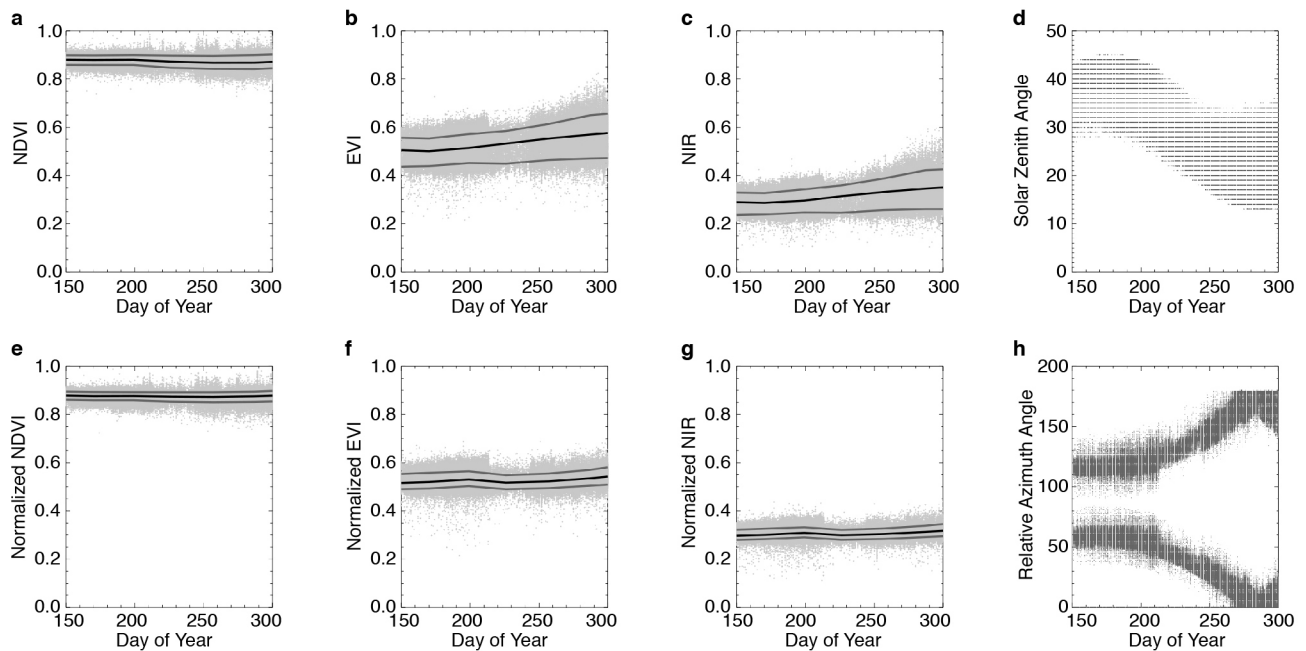
Extended Data Figure 1 | Synthetic Amazon forest developed from field measurements of forest structure and reflectance properties. **a**, Oblique view of the 1 km \times 1 km synthetic forest. **b**, Reflectance (ρ) and transmittance (τ) properties of Amazon leaves (G. P. Asner, unpublished data) and litter⁴¹ at MODIS (red, 620–670 nm; NIR, 841–876 nm) and ICESat (NIR, 1064 nm) wavelengths. **c**, **d**, Parameter values for lidar and optical radiative transfer simulations. Field measurements of ρ_{Litter} differ slightly at optical (860 nm)

and lidar (1064 nm) NIR wavelengths (see **b**). June conditions were used to estimate early dry season (baseline) lidar and optical remote sensing metrics with FLIGHT. **e–g**, Height-diameter relationships (**e**) and crown depth (**f**) and width (**g**) estimates for small trees (< 20-cm diameter at breast height) were derived from allometric relationships for larger trees (> 20-cm diameter at breast height) using regression with a power-law function.



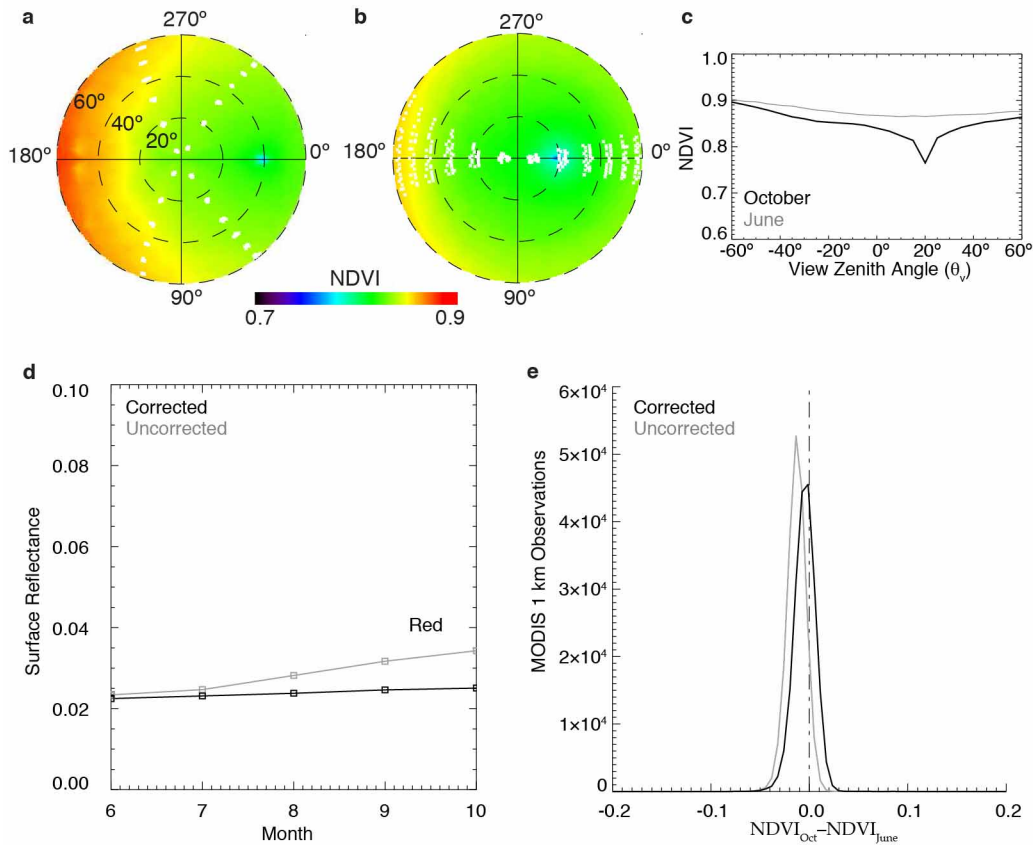
Extended Data Figure 2 | FLIGHT model simulations of GLAS lidar waveforms. **a**, Distribution of modelled waveform centroid relative height (WCRH) for 100 simulated GLAS lidar footprints located at the centre of each 100×100 m grid box of the synthetic Amazon forest. **b**, Nadir view of the

waveform selected for FLIGHT model simulations (blue outline). The dashed vertical line in **a** indicates the June WCRH (0.519) for the footprint location in **b**.



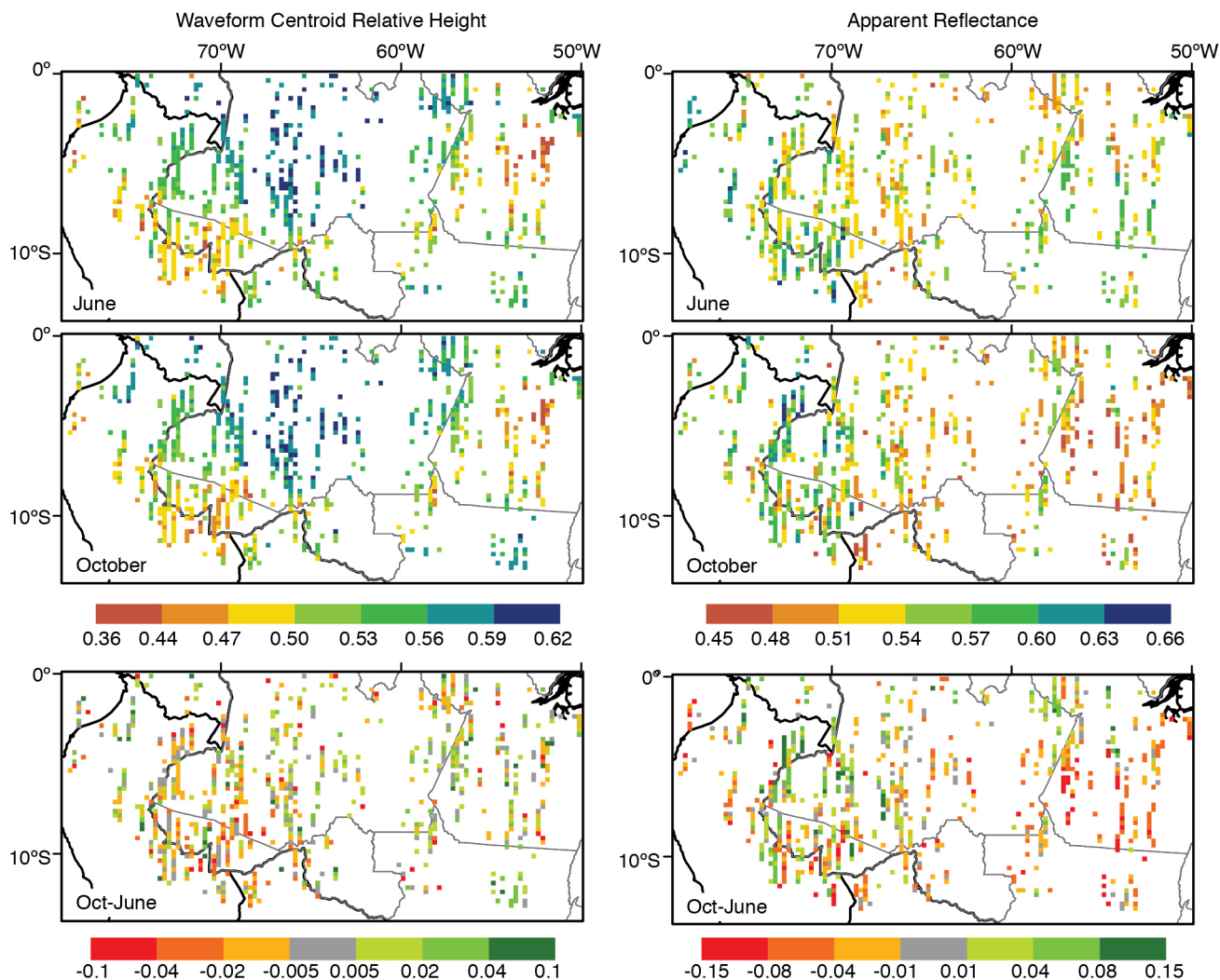
Extended Data Figure 3 | Seasonal variability in MODIS surface reflectance, vegetation indices, and sun-sensor geometry. Seasonal distributions of NDVI, EVI and near-infrared (NIR) reflectance for uncorrected (a–c) and corrected (e–g) MODIS 1 km data from 2003–08, in which corrected MODIS data have been normalized to a consistent sun-sensor geometry. Lines denote

the median (black) and upper and lower quartile median values (grey) for uncorrected and corrected MODIS observations. The decrease in solar zenith angle between day of year 150 and 300 (d) increases the frequency of observations near the principal plane (h, $\phi = 0^\circ$, $\phi = 180^\circ$).



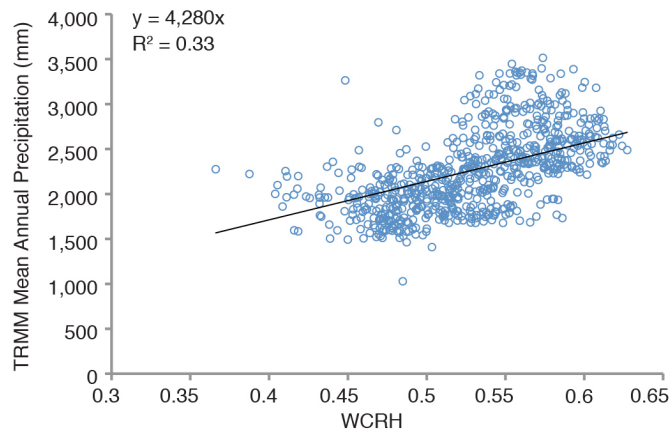
Extended Data Figure 4 | Impact of changing MODIS sun-sensor geometry on NDVI. **a–c,** Seasonal changes in sun-sensor geometry decrease MODIS NDVI over Amazon forest between June and October. As for EVI (see Fig. 1), modelled directional anisotropy of the MODIS NDVI over southern Amazon forests is stronger in June (**a**) than October (**b**), but the realized bidirectional reflectance effect in MODIS observations is greater in October (**b, c**) when the Terra and Aqua MODIS instruments sample in the principal plane ($\phi = 0^\circ$, $\phi = 180^\circ$). For NDVI, the hot spot effect near $\theta_v = 20^\circ$ increases red reflectance, thereby lowering NDVI values at these viewing angles. **d,** Seasonal profile of red

reflectance for uncorrected (grey) and corrected (black) MODIS 1 km data in southern Amazonia. Values indicate the upper quartile median monthly red reflectance from all Terra and Aqua MODIS observations in 2003–08 ($n = 197,651$). **e,** Per-pixel changes in uncorrected (grey) and corrected (black) MODIS NDVI between October and June for all forested areas in southern Amazonia. Small seasonal decreases in MODIS NDVI data were eliminated after normalizing the sun-sensor geometry during the June to October dry season.

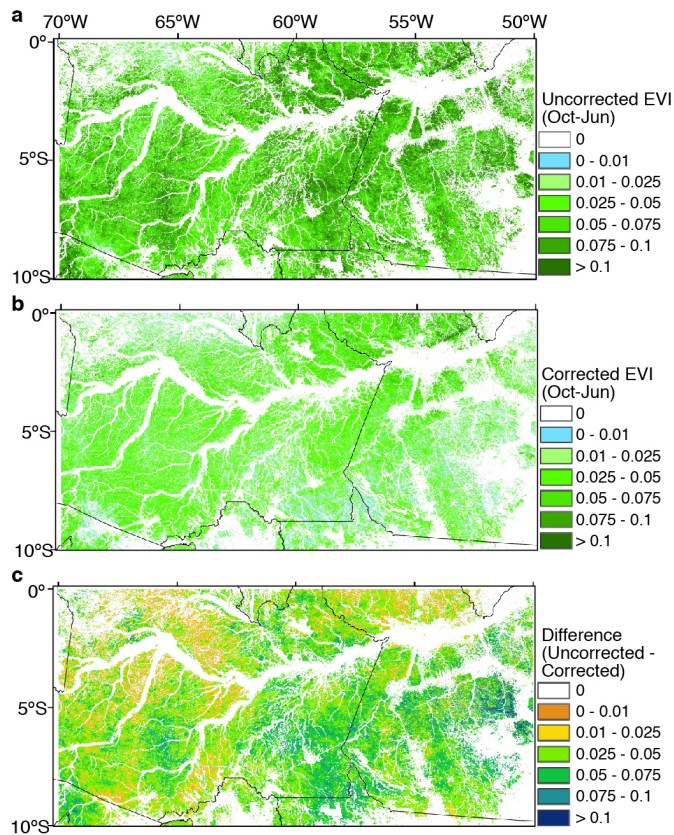


Extended Data Figure 5 | Spatial pattern of seasonal changes in GLAS lidar metrics. a–f, June, October, and difference maps (October minus June) of WCRH (a–c) and 1064 nm apparent reflectance (d–f) for 0.25° grid cells in

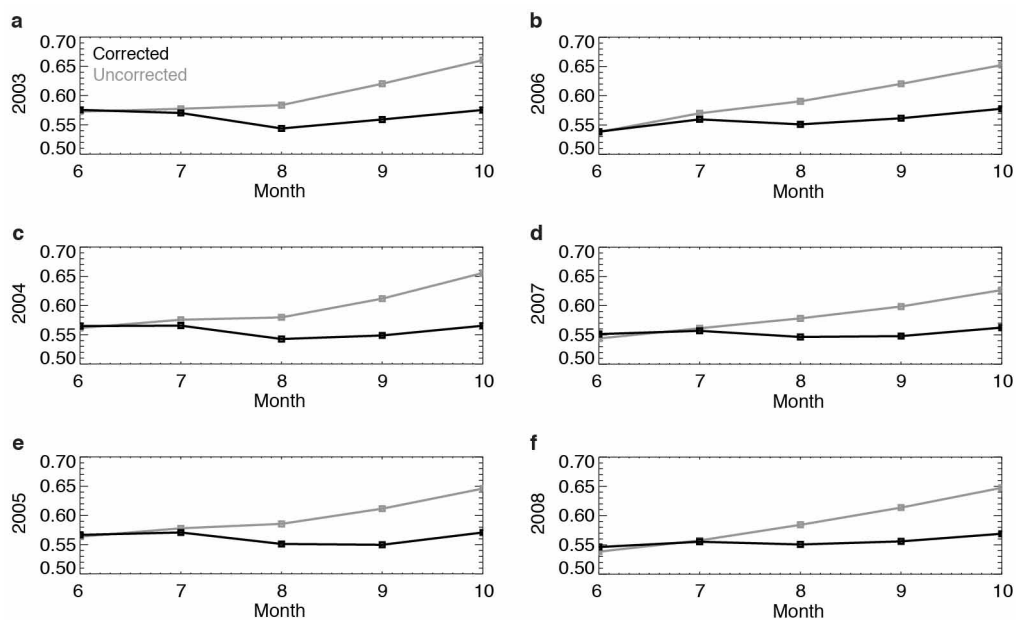
southern Amazonia with ≥ 10 pairs of June and October GLAS footprints. White cells indicate non-forest areas or grid cells with <10 GLAS lidar footprints in one or both months.



Extended Data Figure 6 | Correlation between GLAS lidar WCRH and mean annual precipitation. Values indicate mean June WCRH and mean annual precipitation (September to August, 1997–2009) from the tropical rainfall measurement mission (TRMM) 3B43v6 product for 0.25° cells with ≥ 10 GLAS footprint pairs ($n = 728$, see Extended Data Fig. 5).

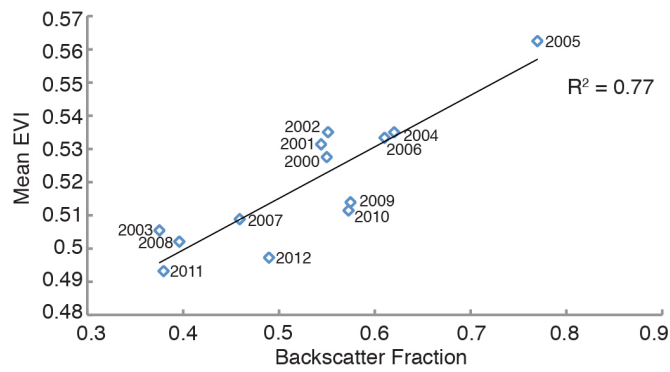


Extended Data Figure 7 | Maps of seasonal amplitude in uncorrected and corrected MODIS EVI data. **a**, Seasonal amplitude in uncorrected MODIS EVI data (October minus June) for two 10° × 10° spatial tiles in southern Amazonia (H11V09, H11V10). **b**, Seasonal amplitude in corrected EVI data (October minus June), using a single BRDF inversion model to normalize changes in sun-sensor geometry for all Amazon forest pixels per tile, per month. Instead of modelling BRDF on a per-pixel basis (see Fig. 3, Extended Data Figs 3, 4 and 8), this map reflects the reduction in the seasonal amplitude of EVI using generic BRDF models derived from all cloud-free MODIS data per tile, per month. **c**, The difference in the seasonal amplitude of MODIS EVI (uncorrected minus corrected) is positive, highlighting how changes in sun-sensor geometry generate the apparent green up phenomenon. White regions indicate non-forest cover types.



Extended Data Figure 8 | Interannual variability in MODIS EVI for southern Amazon forests in 2003–08. Values indicate the upper quartile median monthly EVI for uncorrected (grey) and corrected (black) MODIS

data. Normalization of sun-sensor geometry reduced both seasonal and interannual variability in MODIS EVI over southern Amazon forests.



Extended Data Figure 9 | Contribution of variability in MODIS view geometry to interannual differences in EVI over Amazon forests. EVI and view zenith angle (θ_v) data from the MOD13A1 16-day composite product (data access: <https://lpdaac.usgs.gov>) are shown for spatial tile H11V09 when the Terra MODIS sensor was observing in the principal plane (October, day of year 273–288). The fraction of observations in the backscatter direction ($\theta_v > 0^\circ$) and mean EVI over forests were strongly correlated, highlighting the importance of sun-sensor geometry for interannual variability in uncorrected MODIS EVI data. The drought year of 2005 had the highest fraction of MODIS observations in the backscatter direction of any year during 2000–12.

Extended Data Table 1 | Sensitivity of lidar WCRH and optical vegetation indices (EVI, NDVI) to modelled seasonal changes in forest structure and reflectance properties

Model Parameter	Value	WCRH	EVI	NDVI
LAI	4.5	0.52	0.43	0.81
	5.5	0.55	0.43	0.82
	6.5	0.58	0.43	0.83
Litter (1064 nm)*	0.34	0.52		
	0.44	0.50		
	0.54	0.48		
Litter (860 nm)	0.05		0.38	0.79
	0.30		0.43	0.81
	0.55		0.48	0.83
Leaf†	0.29	0.50		
	0.39	0.52	0.43	0.81
	0.49	0.52	0.54	0.85

Optical data simulations reflect October illumination conditions ($\theta_s = 20^\circ$, $\theta_v = 15^\circ$, $\phi = 0^\circ$).

* Field measurements of ρ_{Litter} differ slightly between lidar (1064 nm) and optical (841–876 nm) NIR ranges (see Extended Data Fig. 1).

† FLIGHT model analyses in the optical domain did not include simulations with $\rho_{\text{Leaf}} = 0.29$.

The genome of a Late Pleistocene human from a Clovis burial site in western Montana

Morten Rasmussen^{1*}, Sarah L. Anzick^{2*}, Michael R. Waters³, Pontus Skoglund⁴, Michael DeGiorgio^{5†}, Thomas W. Stafford Jr^{1,6}, Simon Rasmussen⁷, Ida Moltke^{8,9}, Anders Albrechtsen⁸, Shane M. Doyle¹⁰, G. David Poznik¹¹, Valborg Gudmundsdottir⁷, Rachita Yadav⁷, Anna-Sapfo Malaspinas¹, Samuel Stockton White V¹², Morten E. Allentoft¹, Omar E. Cornejo¹³, Kristiina Tambets¹⁴, Anders Eriksson^{15,16}, Peter D. Heintzman¹⁷, Monika Karmin¹⁴, Thorfinn Sand Korneliusen¹, David J. Meltzer¹⁸, Tracey L. Pierre¹, Jesper Stenderup¹, Lauri Saag¹⁴, Vera M. Warmuth^{15,19}, Margarida C. Lopes¹⁹, Ripan S. Malhi²⁰, Søren Brunak⁷, Thomas Sicheritz-Ponten⁷, Ian Barnes^{17†}, Matthew Collins²¹, Ludovic Orlando¹, Francois Balloux²², Andrea Manica¹⁵, Ramneek Gupta⁷, Mait Metspalu¹⁴, Carlos D. Bustamante^{23,24}, Mattias Jakobsson^{4,25}, Rasmus Nielsen⁵ & Eske Willerslev¹

Clovis, with its distinctive biface, blade and osseous technologies, is the oldest widespread archaeological complex defined in North America, dating from 11,100 to 10,700 ¹⁴C years before present (BP) (13,000 to 12,600 calendar years BP)^{1,2}. Nearly 50 years of archaeological research point to the Clovis complex as having developed south of the North American ice sheets from an ancestral technology³. However, both the origins and the genetic legacy of the people who manufactured Clovis tools remain under debate. It is generally believed that these people ultimately derived from Asia and were directly related to contemporary Native Americans². An alternative, Solutrean, hypothesis posits that the Clovis predecessors emigrated from southwestern Europe during the Last Glacial Maximum⁴. Here we report the genome sequence of a male infant (Anzick-1) recovered from the Anzick burial site in western Montana. The human bones date to 10,705 ± 35 ¹⁴C years BP (approximately 12,707–12,556 calendar years BP) and were directly associated with Clovis tools. We sequenced the genome to an average depth of 14.4× and show that the gene flow from the Siberian Upper Palaeolithic Mal'ta population⁵ into Native American ancestors is also shared by the Anzick-1 individual and thus happened before 12,600 years BP. We also show that the Anzick-1 individual is more closely related to all indigenous American populations than to any other group. Our data are compatible with the hypothesis that Anzick-1 belonged to a population directly ancestral to many contemporary Native Americans. Finally, we find evidence of a deep divergence in Native American populations that predates the Anzick-1 individual.

The only known Clovis burial and associated mortuary assemblage was found in the Americas at the Anzick site, Montana, in 1968 (refs 6–8) (Fig. 1a, b). Here, approximately 100 stone tools and 15 osseous tool fragments (Fig. 1d, e) (at least one made of elk; Supplementary Information section 3.4) that are technologically consistent with artefacts of the Clovis complex^{9,10} were found in direct association with the partial fragmentary

remains of an infant child (Anzick-1). The human remains were found directly below the Clovis artefacts and were covered with red ochre. Bone from the skeleton was directly dated using XAD-collagen to 10,705 ± 35 ¹⁴C years BP (CAMS-80538) or 12,707–12,556 calendar years BP, close to the end of the Clovis time period^{1,7} (Fig. 1c).

Initial genetic screening of the Anzick-1 skeletal remains using PCR coupled with cloning and Sanger sequencing yielded a mitochondrial DNA (mtDNA) haplogroup assignment of D4h3a (Supplementary Information section 3). The D4h3a haplogroup was verified and further characterized in the subsequent shotgun sequencing of Anzick-1 (Supplementary Information section 12 and Extended Data Fig. 2). D4h3a is one of the rare mtDNA lineages specific to Native Americans, is distributed along the Pacific coast in North and South America among contemporary populations¹¹, and is also present in ancient specimens¹². Its current distribution has been interpreted as evidence for an early coastal migration route¹¹. Our findings of this mtDNA haplogroup inland in the oldest skeleton from the Americas to be mtDNA-typed to date question such interpretation and underscore the view that current distributions of genetic markers are not necessarily indicative of the movement or distribution of peoples in the past³. The Anzick-1 D4h3a does not carry any of the polymorphisms that define the several subgroups of the haplogroup¹³ and is thus placed at the root of D4h3a. Our finding suggests that the origin of the D4h3a branch is likely to be at the upper bound of the previously obtained estimate of 13,000 ± 2,600 calendar years BP¹⁴, or possibly even older.

Shotgun sequencing of the Anzick-1 DNA revealed a highly variable endogenous human DNA content across different extracts, even those recovered from the same bone fragment (0.5–28.2%; Supplementary Information section 4 and Extended Data Table 3). This suggests a considerable influence of microenvironments on long-term DNA preservation across the Anzick-1 bone fragments¹⁵. We obtained an average sequencing depth of 14.4× coverage. On the basis of the observed DNA fragment

¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark. ²Anzick Family, 31 Old Clyde Park Road, Livingston, Montana 59047, USA. ³Center for the Study of the First Americans, Departments of Anthropology and Geography, Texas A&M University, 4352 TAMU, College Station, Texas 77843-4352, USA.

⁴Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden. ⁵Department of Integrative Biology, University of California, Berkeley, 4134 Valley Life Sciences Building, Berkeley, California 94720, USA. ⁶AMS 14C Dating Centre, Department of Physics & Astronomy, University of Aarhus, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ⁷Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet 208, Kgs. Lyngby DK-2800, Denmark. ⁸The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark. ⁹Department of Human Genetics, University of Chicago, 920 E. 58th Street, CLSC 4th floor, Chicago, Illinois 60637, USA. ¹⁰Education Department, Montana State University, Box 5103, Bozeman, Montana 59717, USA. ¹¹Program in Biomedical Informatics and Department of Statistics, Stanford University, Stanford, California 94305, USA. ¹²Anthropology Department, PhD Program, University of Montana, 4100 Mullan Road, no. 217, Missoula, Montana 59808, USA. ¹³School of Biological Sciences, Washington State University, PO Box 644236, Eastlick Hall 395, Pullman, Washington 99164, USA. ¹⁴Department of Evolutionary Biology, Estonian Biocentre and University of Tartu, Riia 23b, 51010 Tartu, Estonia. ¹⁵Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ¹⁶Integrative Systems Biology Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia. ¹⁷School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. ¹⁸Department of Anthropology, Southern Methodist University, Dallas, Texas 75275, USA. ¹⁹Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK. ²⁰Department of Anthropology and Institute for Genomic Biology, University of Illinois Urbana-Champaign, 209F Davenport Hall, 607 Matthews Avenue, Urbana, Illinois 61801, USA. ²¹BioArCh, Departments of Biology, Archaeology and Chemistry, University of York, Wentworth Way, York YO10 5DD, UK. ²²MRC Centre for Outbreak, Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, Imperial College Faculty of Medicine, London W2 1PG, UK. ²³Department of Genetics, School of Medicine, Stanford University, Littlefield Center, Stanford, California 94305, USA. ²⁴Center for Evolutionary and Human Genomics, Stanford University, Littlefield Center, Stanford, California 94305, USA. ²⁵Science for Life Laboratory, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden. †Present addresses: Earth Sciences Department, Natural History Museum, Cromwell Road, London SW7 5BD, UK (I.B.); Department of Biology, Pennsylvania State University, 502 Wartik Laboratory, University Park, Pennsylvania 16802, USA (M.D.).

*These authors contributed equally to this work.

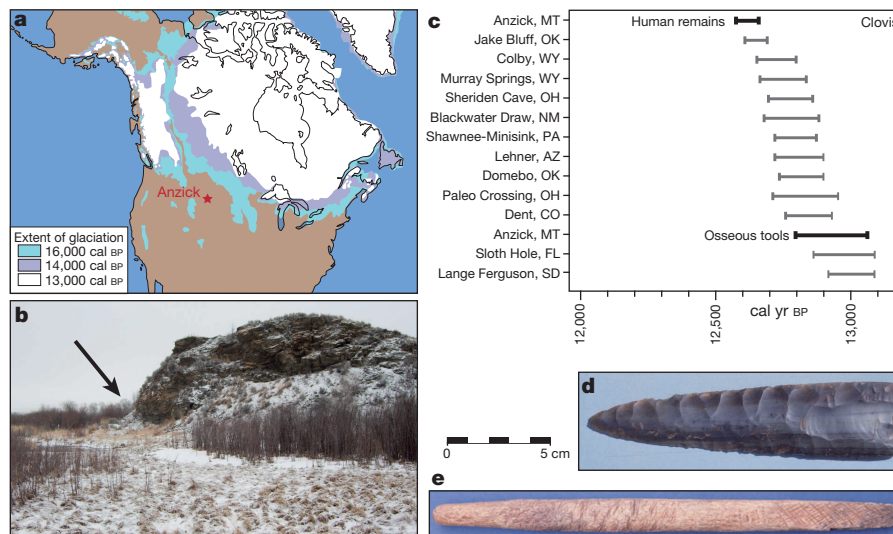


Figure 1 | Geographic and ^{14}C dating overview, and examples of artefacts from the site. a, Location of the Anzick site relative to continental glacial positions from 16,000 to 13,000 calendar years BP (cal BP). **b**, Photograph of the

length distribution, we computed molecular half-life estimates¹⁶ for a 100-base pair fragment to be 4,852 and 5,137 years for nuclear DNA and mtDNA, respectively. These are consistent with predictions based on estimated burial temperature and geographical location (Supplementary Information section 7). Nucleotide misincorporation patterns were consistent with increased post-mortem cytosine deamination at DNA fragment termini¹⁷ (Supplementary Information section 8). The mtDNA data provided a contamination estimate with a 95% credibility interval of 0.15–0.38%¹⁸ (Supplementary Information section 9). As the Anzick-1 individual was found to be male, we were also able to quantify contamination levels on the basis of heterozygosity levels seen on the X chromosome¹⁹, which resulted in similarly low contamination estimates (<1.2%, Supplementary Information section 11). Characterization of the error rate yielded an overall estimate of 0.84% (Supplementary Information section 10). Per-library error rates as well as nucleotide-type-specific error rates are shown in Extended Data Fig. 1c; as expected, the error estimates are driven by deamination-related DNA-damage errors.

We determined the Y-chromosome haplogroup to be Q-L54*(xM3) and, along with 15 previously analysed Y-chromosome sequences²⁰, we constructed a tree to illustrate the phylogenetic context within haplogroup Q (Supplementary Information section 13 and Extended Data Fig. 2). Confining our analyses to transversion single nucleotide polymorphisms (SNPs), we leveraged the date of Anzick-1 to estimate a divergence time between haplogroups Q-L54*(xM3) and Q-M3, two of the major founding Y-chromosome lineages of the Americas, of approximately 16,900 years ago (95% confidence interval: 13,000–19,700; Supplementary Information section 13).

We assessed the genome-wide genetic affinity of the Anzick-1 individual to 143 contemporary non-African human populations by computing outgroup f_3 -statistics²¹, which are informative on the amount of shared genetic drift between an individual and other populations. The data set included 52 Native American populations, for which genomic segments derived from recent European and African admixture have been excluded²². We found that the Anzick-1 individual showed a statistically significant closer affinity to all 52 Native American groups than to any extant Eurasian population (Fig. 2). The same conclusion was reached from admixture clustering analyses (Supplementary Information section 14 and Extended Data Fig. 3), and when using D -statistics based on genomic sequence data (Supplementary Information section 16).

Interestingly, the Anzick-1 individual showed less shared genetic history with seven Northern Native Americans from Canada and the Arctic, including three Northern Amerind-speaking groups ('NA' group),

Anzick site. The site is located at the base of the slope at the far left (arrow). **c**, Age of the human remains and osseous tools relative to other Clovis sites. **d**, Clovis fluted projectile point from the site. **e**, Clovis osseous rod from the site.

and the Central American Yaqui (Supplementary Information section 15, Fig. 2), than with 44 Native populations from Central and South America ('SA' group) (Fig. 2). This finding was further supported by allele-frequency-based D -statistics, which strongly reject the hypothesis of a population tree with topology (Anzick-1, (NA, SA)). By contrast, a tree assuming a branching pattern of (NA, (Anzick-1, SA)) was not rejected (Supplementary Information section 15 and Fig. 3).

We investigated two possible scenarios that could explain this pattern: (1) a basal diversification of the NA and SA lineages that predates the Clovis period, with the Anzick-1 individual belonging to the SA lineage (Fig. 3c); and (2) the Anzick-1 individual is basal to both lineages, but the lineages diverged at a time close to approximately 12,600 years ago, and the NA group received subsequent gene-flow from other, perhaps Siberian, populations (Fig. 3b). This latter hypothesis of additional pulses of gene flow has been suggested for the four Na-Dene- and Eskimo-Aleut-speaking groups by a previous study, whereas the three NA groups and Yaqui have been found to carry exclusively 'First American' ancestry²².

We analysed 19 different published Siberian populations and found that allele-frequency-based D -statistics in all cases, with the exclusion of Naukan, were compatible with a diversification pattern of (Siberian, (NA, SA)), with no evidence for gene flow into the NA group (Supplementary Information section 15), a pattern that agrees with previously published results²². This suggests that model (1) is more likely, and that the structure between Native American lineages predates the Anzick-1 individual and thus appears to go back to pre-Clovis times.

We used outgroup f_3 -statistics to evaluate the shared genetic history between all Native American populations and the Anzick-1 genome, the 24,000-year-old human sample from Mal'ta, Siberia⁵ and the 4,000-year-old Saqqaq Palaeo-Eskimo sample from Greenland¹⁹. We again found a closer relationship between Anzick-1 and all Native Americans (Supplementary Information section 15 and Extended Data Fig. 5). Together with the fact that Anzick-1 shows the same relative affinity to western and eastern Eurasians, this suggests that the gene flow from the Mal'ta lineage into Native Americans happened before the NA and SA groups diverged.

Next, we addressed the relationship of the Anzick-1 genome to whole-genome sequences from contemporary humans, including two novel genomes from Karitiana and Mayan individuals, and from the ancient Saqqaq sample¹⁹. To explore the question of direct ancestry, we developed a novel method to assess the amount of genetic drift experienced by each of a pair of populations after their divergence, without making any assumptions about demography (Supplementary Information section

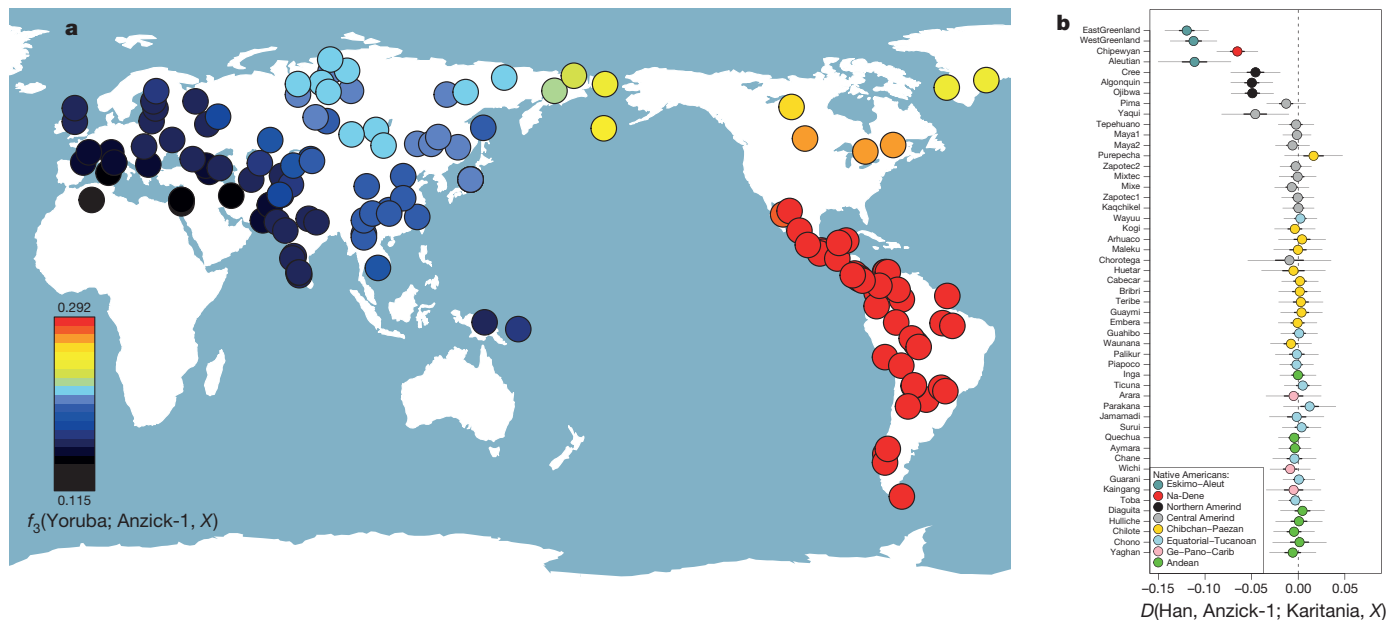


Figure 2 | Genetic affinity of Anzick-1. **a**, Anzick-1 is most closely related to Native Americans. Heat map representing estimated outgroup f_3 -statistics for shared genetic history between the Anzick-1 individual and each of 143 contemporary human populations outside sub-Saharan Africa. **b**, Anzick-1 is less closely related to Northern Native American populations and a Yaqui

individual than to Central and South Native Americans such as the Brazilian Karitiana. We computed D -statistics of the form $D(\text{Han, Anzick-1; Karitania, } X)$ to test the hypothesis that a second Native American population X is as closely related to Anzick-1 as the South American Karitiana is. Thick and thin whiskers represent 1 and 3 standard errors, respectively.

17). If the Anzick-1 individual belonged to a population directly ancestral to a sampled contemporary population, the amount of genetic drift on the Anzick-1 branch should be zero, as it would have experienced no genetic drift since the 'divergence' with the sampled population. We find that the data are compatible with the Anzick-1 individual belonging to a population that is directly ancestral to the two South American Karitiana samples, as is the case for the Mayan, after masking the latter for recent European admixture (Fig. 4a, b). By contrast, there is a positive amount of drift on both lineages when comparing Anzick-1 to Saqqaq, Europeans, Asians and Pacific Islanders. This shows that the Anzick-1 sample has diverged from populations from outside the Americas. Furthermore, in agreement with the SNP analyses, *TreeMix*²³ (Fig. 4c

and Supplementary Information section 18) and outgroup f_3 -analyses^{5,21} (Extended Data Figs 5, 6), the Anzick-1 sample is genetically more closely related to Central and South Americans than to any other populations, including the Saqqaq individual from Greenland. After masking the Mayan genome for recent European admixture, *TreeMix* places the Anzick-1 individual in a position in the tree compatible with the hypothesis that it is ancestral to both Mayan and Karitiana, with Anzick-1 exhibiting virtually no drift on its branch since its divergence from other populations (Fig. 4c).

We conclude that the male infant, buried approximately 12,600 years ago with ochre-covered Clovis artefacts at the Anzick site, belonged to a meta-population from which many contemporary Native Americans are

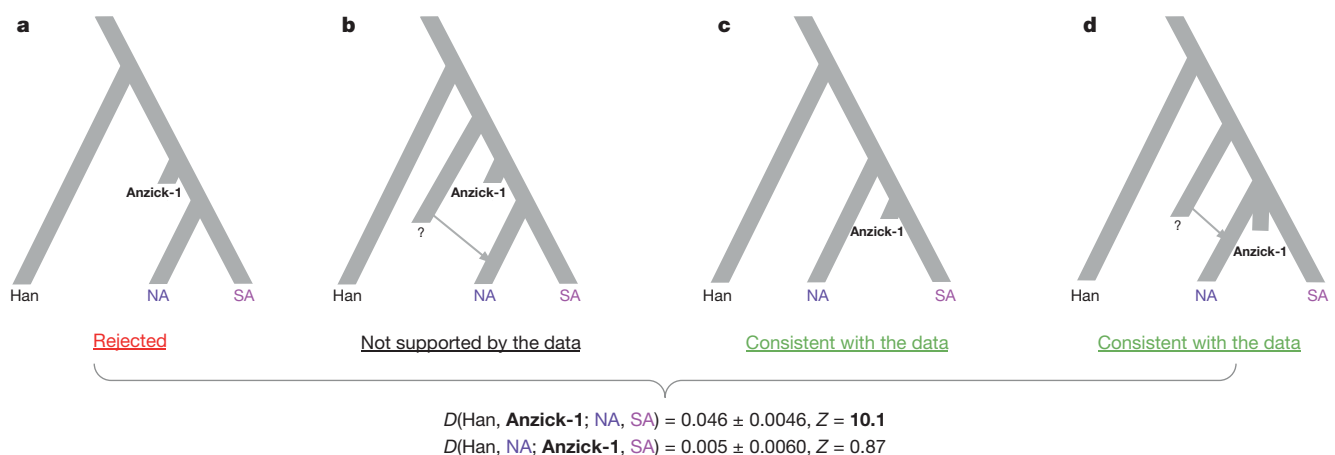


Figure 3 | Simplified schematic of genetic models. Alternative models of the population history behind the closer shared ancestry of the Anzick-1 individual to Central and Southern American (SA) populations than Northern Native American (NA) populations; see main text for further definition of populations. We find that the data are consistent with a simple tree-like model in which NA populations are historically basal to Anzick-1 and SA. We base this conclusion on two D -tests conducted on the Anzick-1 individual, NA and SA. We used Han Chinese as outgroup. **a**, We first tested the hypothesis that Anzick-1 is basal to both NA and SA populations using $D(\text{Han, Anzick-1; NA, SA})$. As in

the results for each pairwise comparison between SA and NA populations (Extended Data Fig. 4), this hypothesis is rejected. **b**, Next, we tested $D(\text{Han, NA; Anzick-1, SA})$; if NA populations were a mixture of post-Anzick-1 and pre-Anzick-1 ancestry, we would expect to reject this topology. **c**, We found that a topology with NA populations basal to Anzick-1 and SA populations is consistent with the data. **d**, However, another alternative is that the Anzick-1 individual is from the time of the last common ancestral population of the Northern and Southern lineage, after which the Northern lineage received gene flow from a more basal lineage.

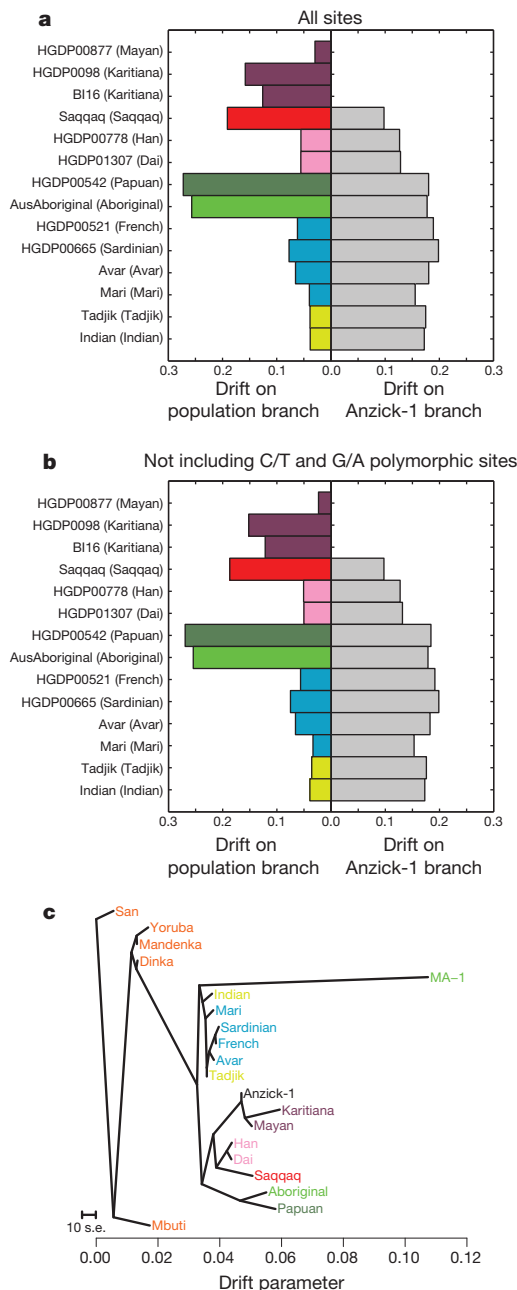


Figure 4 | Estimated drift and maximum likelihood tree. Site patterns and drift estimates for non-African populations versus the Anzick-1 sample. **a**, Data based on conditioning on African polymorphism, non-African populations in colour on the left, Anzick-1 in grey on the right. **b**, Data based on conditioning on African polymorphism and removing sites where a C and T or G and A were observed. **c**, Maximum likelihood tree generated by *TreeMix* using the whole-genome sequencing data, with the Mayan genome masked for European ancestry. HGDP, Human Genome Diversity Project.

descended and is closely related to all indigenous American populations. As such, contemporary Native Americans are effectively direct descendants of the people who made and used Clovis tools and buried this child. In agreement with previous archaeological and genetic studies^{24,25}, our genome analysis refutes the possibility that Clovis originated via a European (Solutrean) migration to the Americas⁴. Furthermore, the Anzick-1 findings do not support the hypothesis, based on cranial data, that American populations around the time of Clovis were subsequently assimilated by more recent migrants who were the ancestors of contemporary Native Americans^{26,27}. The Anzick-1 data thus serves to unify the genetic and archaeological records of early North America, it is consistent

with a human occupation of the Americas a few thousand years before Clovis^{2,3,28}, and demonstrates that contemporary Native Americans are descendants of the first people to settle successfully in the Americas. Our results are also consistent with previous models derived from mtDNA, which imply that Native American populations primarily derive from a single-source population, but that there was a secondary movement into northern North America²⁹. However, several different scenarios are compatible with an early divergence of the NA and SA groups and analyses of more ancient human remains are needed to further test the findings and interpretations from this single individual and to elucidate the complex colonization history of the Arctic and North American populations.

METHODS SUMMARY

We sampled bone fragments from the Anzick-1 skeleton, from the Anzick site in Montana, for ancient DNA and ¹⁴C dating analyses (Supplementary Information sections 1, 2). From DNA extracts we generated Illumina sequencing libraries, which were sequenced on the Illumina HiSeq platform (Supplementary Information sections 3, 4). We verified the authenticity of the ancient DNA through damage patterns and decay rates (Supplementary Information sections 7, 8), and estimated contamination using analyses of both mtDNA and nuclear DNA (Supplementary Information sections 9, 10). To investigate the affinity of the Anzick-1 individual to worldwide modern-day populations we computed outgroup f_3 -statistics^{5,21} using a genome-wide SNP data set of 143 groups collected from multiple studies (Supplementary Information section 15). To test historical models relating the Anzick-1 individual to modern-day Native Americans from a Northern group (Cree, Ojibwa and Algonquin) and a Southern group (44 Central and South American groups), we computed D -statistics²¹ using the full allele frequency information (Supplementary Information section 15). To test whether the Anzick-1 individual could have belonged to a population ancestral to some of the modern populations, we developed a new maximum likelihood method for estimating the length of the branches of a drift tree with two populations (leaves in the tree) represented by two genomes. The method fits a model with the following parameters: the probability of pairwise coalescence within population 1 (c_1), the probability of pairwise coalescence within population 2 (c_2), and the probabilities of genotype patterns in the ancestral population, to the observed counts of joint genotype patterns of the two genomes. These parameters are uniquely determined by the observed genotype configurations. Letting the Anzick-1 individual be represented by population 1, the null hypothesis that it belongs to an ancestral population of a specific modern population (population 2) then formally corresponds to a test of the hypothesis $c_1 = 0$ (Supplementary Information section 17).

Statement regarding legal and ethical issues. Acknowledging the complex ethical and legal issues surrounding the research of Native American human remains in the United States, we have striven not only to comply with federal and state law but also to proactively consult with Native American tribes. The Anzick burial site was discovered on private land and the remains recovered have not been in control of a federally funded museum or federal agency, and thus the Native American Graves Protection and Repatriation Act (NAGPRA) does not apply. Under Montana state law, unmarked human burials are not considered abandoned. Advice provided to the project by members of the Montana State Burial Board, however, confirmed that as no claimant has made a request for the remains, the human remains from the Anzick burial site remain under the control of the landowners, the Anzick family. However, to ensure that Native American concerns were addressed, we have informed nine Native American groups with reservations in the surrounding area of the Anzick site about our work. E.W. and S.M.D. visited in person the cultural representatives of the Northern Cheyenne, Crow, Black Feet and Salish and Kootenai tribes. S.M.D. was in direct contact with the Lakota, Rocky Boys, Assiniboine, Gros Ventre and Chippewa cultural representatives. We received no objections to our research from these groups; however, many of the tribal representatives requested reburial of the remains. The Anzick family is working towards reburial of the human remains from the site.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 November 2013; accepted 14 January 2014.

- Waters, M. R. & Stafford, T. W. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* **315**, 1122–1126 (2007).
- Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**, 1497–1502 (2008).
- Meltzer, D. J. *First Peoples in a New World* (Univ. California Press, 2009).

4. Stanford, D. J. & Bradley, B. A. *Across Atlantic Ice: the Origin of America's Clovis Culture* (Univ. California Press, 2012).
5. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
6. Lahren, L. & Bonnicksen, R. Bone foreshafts from a Clovis burial in southwestern Montana. *Science* **186**, 147–150 (1974).
7. Owsley, D. W. & Hunt, D. Clovis and early Archaic crania from the Anzick site (24PA506), Park County, Montana. *Plains Anthropol.* **46**, 115–124 (2001).
8. Lahren, L. A. *Homeland* (Cayuse Press, 2006).
9. Bradley, B. A., Collins, M. B. & Hemmings, A. *Clovis Technology* (Archaeological Series 17, 2010).
10. Morrow, J. E. & Fiedel, S. J. in *Paleoindian Archaeology: a Hemispheric Perspective* (eds Morrow, J. E. & Fiedel, S. J.) 123–138 (Univ. Press of Florida, 2006).
11. Perego, U. A. *et al.* Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **19**, 1–8 (2009).
12. Kemp, B. M. *et al.* Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am. J. Phys. Anthropol.* **132**, 605–621 (2007).
13. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
14. Behar, D. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
15. Pruvost, M. *et al.* Freshly excavated fossil bones are best for amplification of ancient DNA. *Proc. Natl Acad. Sci. USA* **104**, 739–744 (2007).
16. Allentoft, M. E. *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. Lond. B* **279**, 4724–4733 (2012).
17. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
18. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
19. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
20. Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562–565 (2013).
21. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
22. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
23. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
24. Straus, L. G., Meltzer, D. J. & Goebel, T. Ice Age Atlantis? Exploring the Solutrean-Clovis 'connection'. *World Archaeol.* **37**, 507–532 (2005).
25. Fagundes, N. J. R. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* **82**, 583–592 (2008).
26. Jantz, R. L. & Owsley, D. W. Variation among early North American crania. *Am. J. Phys. Anthropol.* **114**, 146–155 (2001).
27. Steele, D. G. & Powell, J. F. in *Who Were the First Americans: Proceedings of the 58th Annual Biology Colloquium* 97–126 (Center for the Study of the First Americans, Oregon State Univ., 1999).
28. Dillehay, T. D. *et al.* Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* **320**, 784–786 (2008).
29. Kemp, B. M. & Schurr, T. G. in *Human Variation in the Americas* (ed. Auerbach, B. M.) 12–50 (Southern Illinois Univ., 2010).

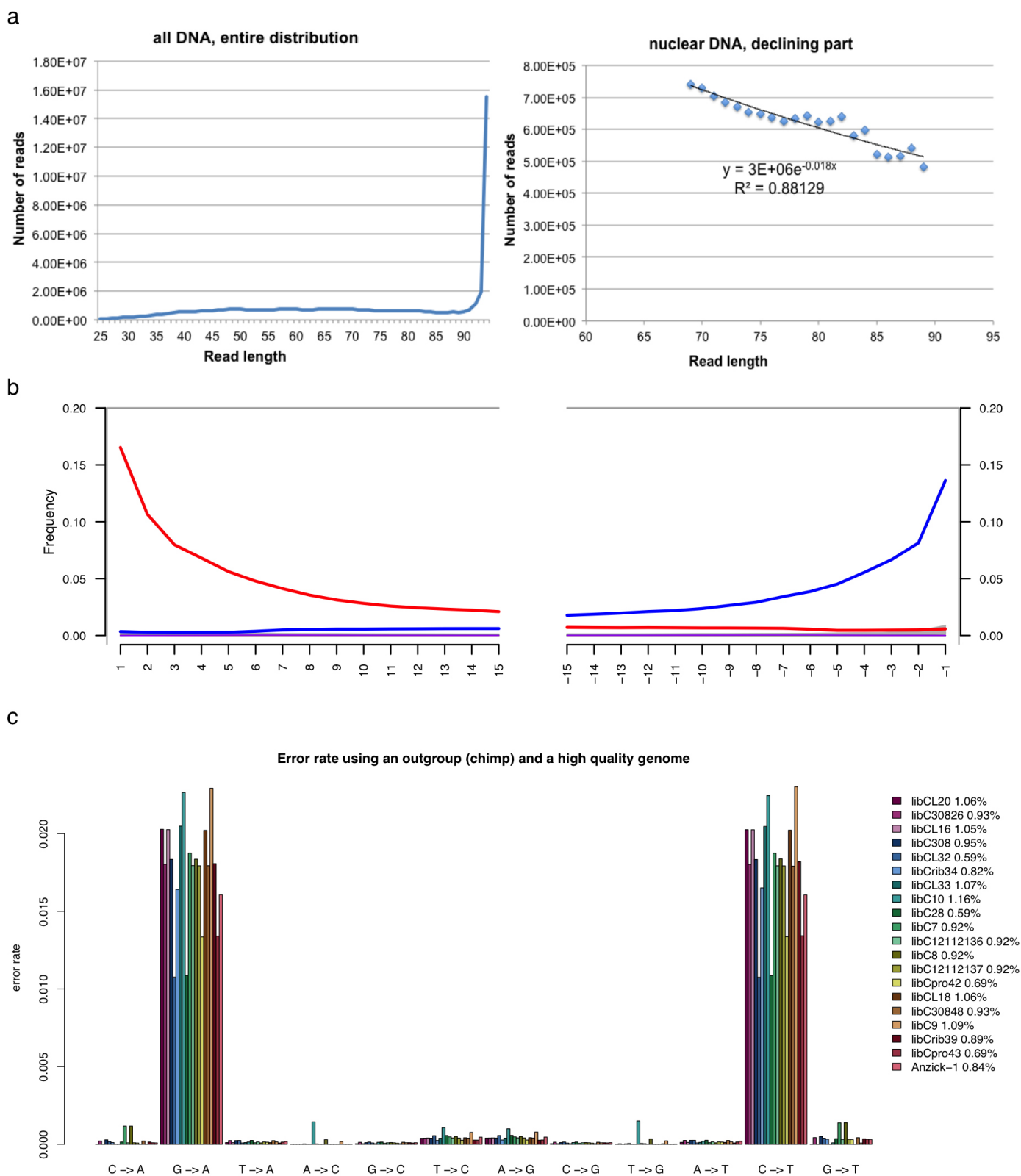
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Danish National High-throughput DNA Sequencing Centre for help with sequencing. B. Henn and J. Kidd for assistance with Human

Genome Diversity Project data, J. Keene for help with illustrations, M. Li, P. L. F. Johnson and M. Stoneking for help with the mtDNA analysis, L. A. Lahren for input to the site description and for establishing contact with the Native American groups, and J. E. Morrow, S. Fiedel and E. Lorenzen for comments on the manuscript. GeoGenetics were supported by the Lundbeck Foundation and the Danish National Research Foundation (DNRF94). M.D. was supported by the US National Science Foundation (grant DBI-1103639). A.-S.M. was supported by the Swiss National Science foundation. G.D.P. was supported by National Science Foundation (NSF) graduate research fellowship DGE-1147470. M.M., M.K., K.T. and L.S. were supported by the European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre and University of Tartu, Estonian Basic Research (grant SF0270177As08) and Estonian Science Foundation (grant 8973). Computations in Uppsala were performed on resources provided by SNIC-UPPMAX (project b2012063) and in Tartu using the High Performance Computing Centre of the University of Tartu. A.E., V.M.W., M.C.L., F.B. and A.M. were supported by the Biotechnology and Biological Sciences Research Council (grant P25032 and BB/H005854/1). We thank the North Star Archaeological Research Program, Center for the Study of the First Americans, Texas A&M University, E. Hill, and Stafford Research, Inc. for funding some of the project.

Author Contributions E.W., S.L.A., M.R.W. and T.W.S. conceived the project. E.W. headed the project. E.W. and M.R. designed the research project setup. R.N. supervised the bioinformatical and population genetic analyses with input from M.J. S.M.D., R.S.M. and T.L.P. helped with ethics and contact to local communities. S.L.A. and M.R.W. provided access to the Anzick-1 sample and the archaeological context, with input from S.S.W. and D.J.M. T.W.S. performed AMS ^{14}C dating, stable isotope analyses and provided geochemical and geochronological assessments. S.L.A. and J.S. performed initial mtDNA screening experiments. Elk extracts were processed by P.D.H. and I.B. Ancient DNA extractions and library constructs for shotgun sequencing and preparation for sequencing was done by M.R. O.E.C. prepared the two modern genomes. M.R. and S.R. did initial bioinformatics and mapping of the ancient sample. Mapping of modern samples, and genotype calls was done by S.R., with input from T.S.K., A.E., V.M.W. and M.C.L. S.R., T.S.-P. and S.B. provided super computing resources. O.E.C. and S.R. did phasing and ancestry painting, with input from A.E., V.M.W. and M.C.L. M.E.A. and M.C. did half-life estimates, with input on geology from T.W.S. DNA damage patterns were done by M.R. and L.O. mtDNA consensus and damage estimate was done by A.-S.M. I.M. and A.A. performed the X-chromosome contamination estimates, error rate estimates and *D*-statistic analyses on genomic sequence data. G.D.P. conducted Y-chromosome analysis with input from C.D.B. M.M. did ADMIXTURE analysis. K.T., M.K. and M.M. did mtDNA characterization. P.S. did f_3 -statistics on SNP array data and tested Native American population models using *D*-statistics. M.D. performed *TreeMix* analysis and genome-wide f_3 -statistics. R.N. and M.D. developed and implemented the ancestry test. M.R., S.L.A., M.R.W., P.S., M.D., R.N. and E.W. wrote most of the manuscript with input from T.W.S., M.E.A., A.-S.M., S.R., I.M., A.A., G.D.P., K.T., M.M., R.G., V.G., R.Y., P.D.H., O.E.C., M.C., F.B., A.M., L.S. and the remaining authors.

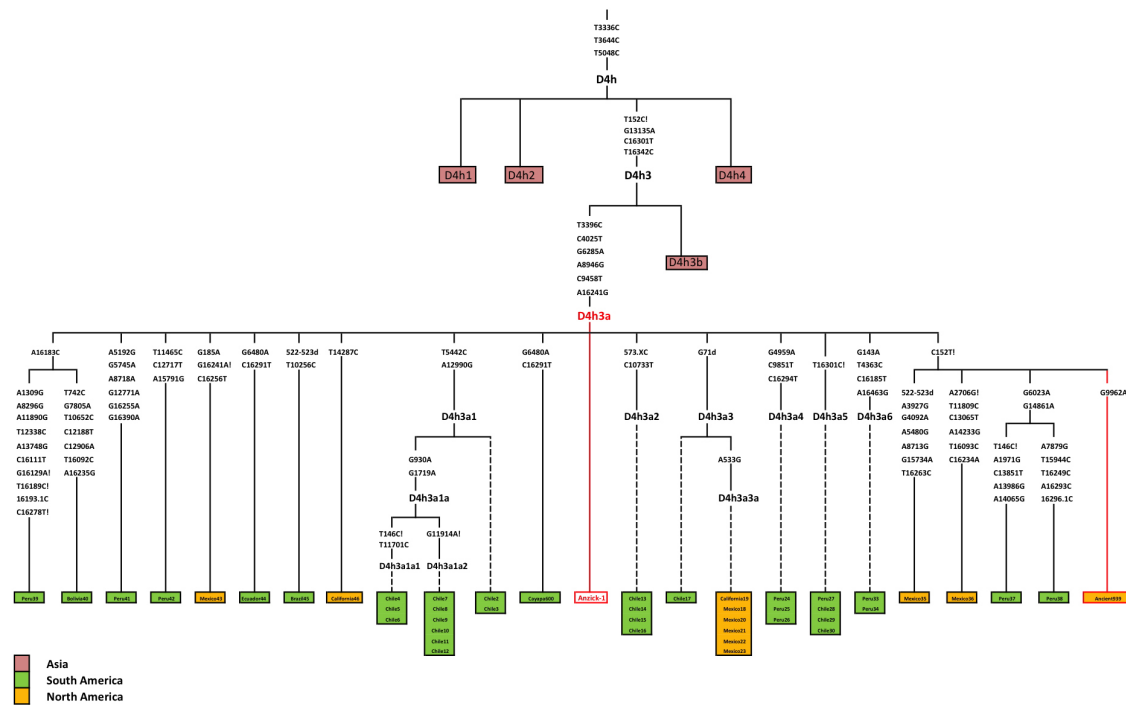
Author Information Sequence data (fastq files) for Anzick-1 is available for download through NCBI SRA accession number SRX381032. Additionally, alignments and genotype calls are available for download at <http://www.cbs.dtu.dk/suppl/clovis/>. Raw reads (fastq files) and alignments (BAM files) for the two modern genomes sequenced in this study are available for demographic research under data access agreement with E.W. The *Cervus elaphus* sequences are available under GenBank accessions KF906070, KF906071 and KF906072. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk).



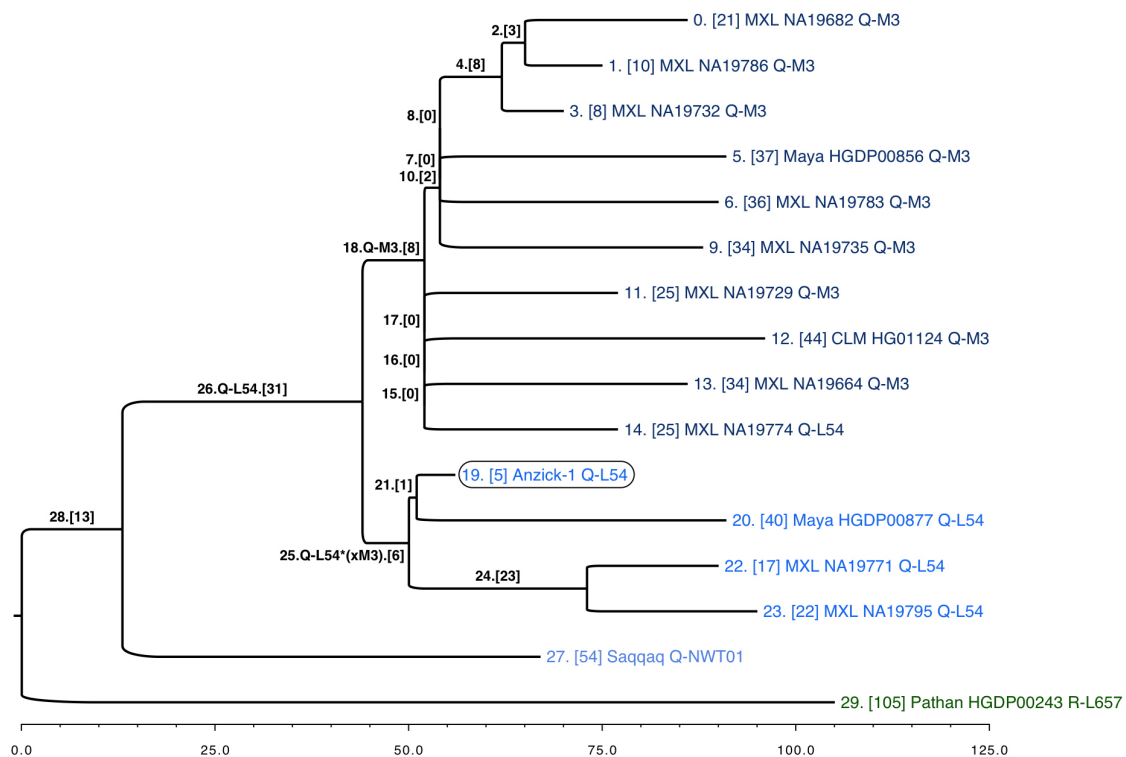
Extended Data Figure 1 | DNA fragmentation, damage and type-specific error. **a**, Left, fragment length distribution of the Anzick-1 DNA sequences mapping to a human reference genome. The maximum read length with the applied chemistry on the HiSeq Illumina platform is 94 base pairs (100 – 6 base pair index read), hence the large peak at this length simply represents the entire tail of the distribution. Right, the declining part of the distribution for the

nuclear DNA, and the fit to an exponential model. The decay constant (λ) is estimated to 0.018. **b**, Damage patterns for the Anzick-1 individual in a random 0.5% subset of all mapped reads. Mismatch frequency relative to the reference as function of read position, C to T in red and G to A in blue. **c**, Type-specific error rates for the Anzick-1 sample and the individual libraries. Estimates of overall error rates are given on the right.

a



b

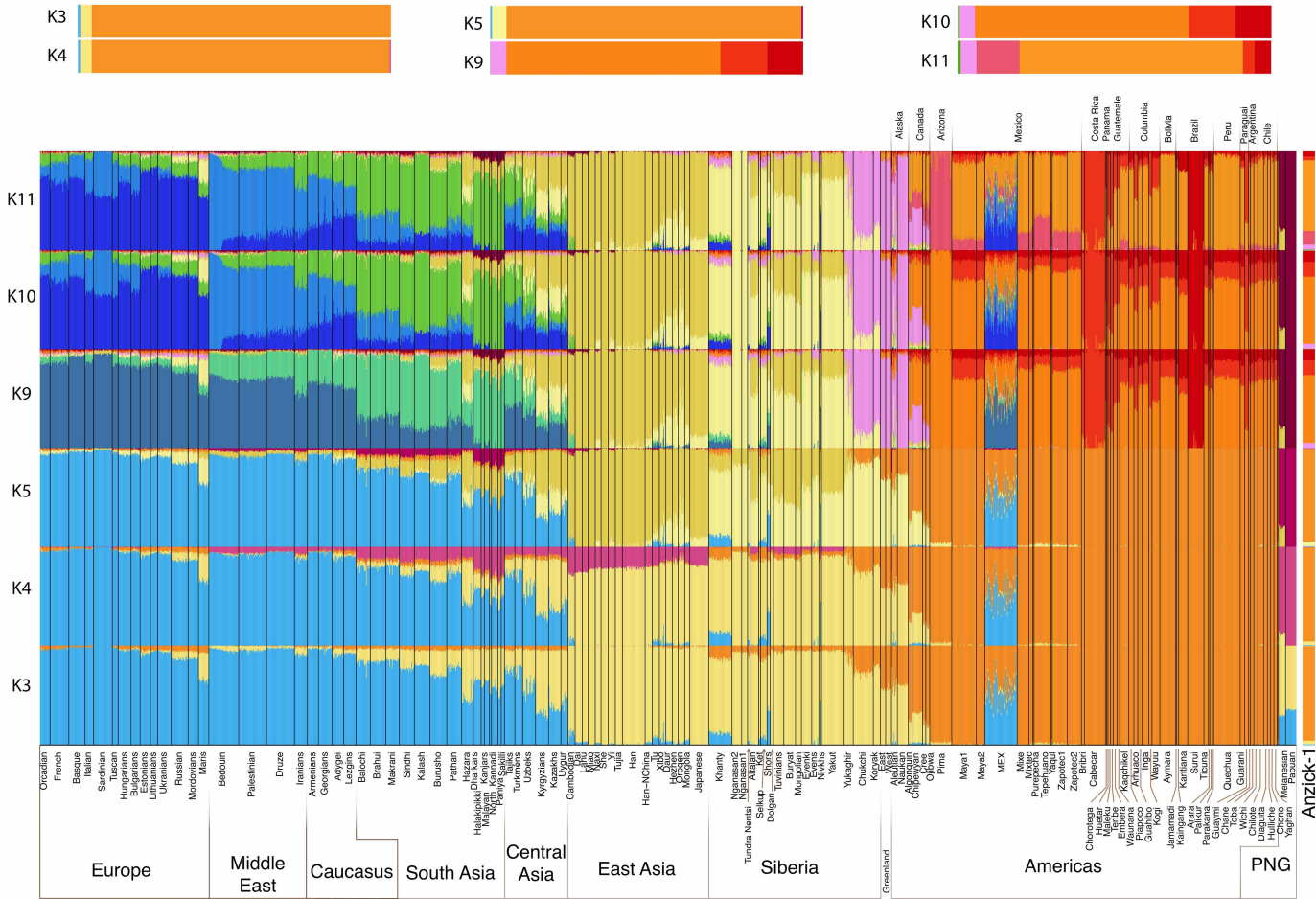


Extended Data Figure 2 | mtDNA and Y-chromosome subtrees.

a, Schematic phylogenetic tree of mtDNA haplogroup D4h3 and its sub-branch D4h3a. Mutations from the root of haplogroup D4h are specified only for haplogroup D4h3a lineage; diagnostic mutations are shown only for defined sub-branches on solid lines. The haplotypes of Anzick-1, identical with the root haplotype of D4h3a, and an ancient full sequence from the northwestern coast of North America (Ancient939), are indicated in red. Insertions are indicated with '.' followed by a number of inserted nucleotides (X if not specified), deletions are indicated with 'd' and back mutations to ancestral state with '*'. The geographical spread of sub-branches of haplogroup D4h is shown with different colours specified in figure legend. **b**, Placement of Anzick-1

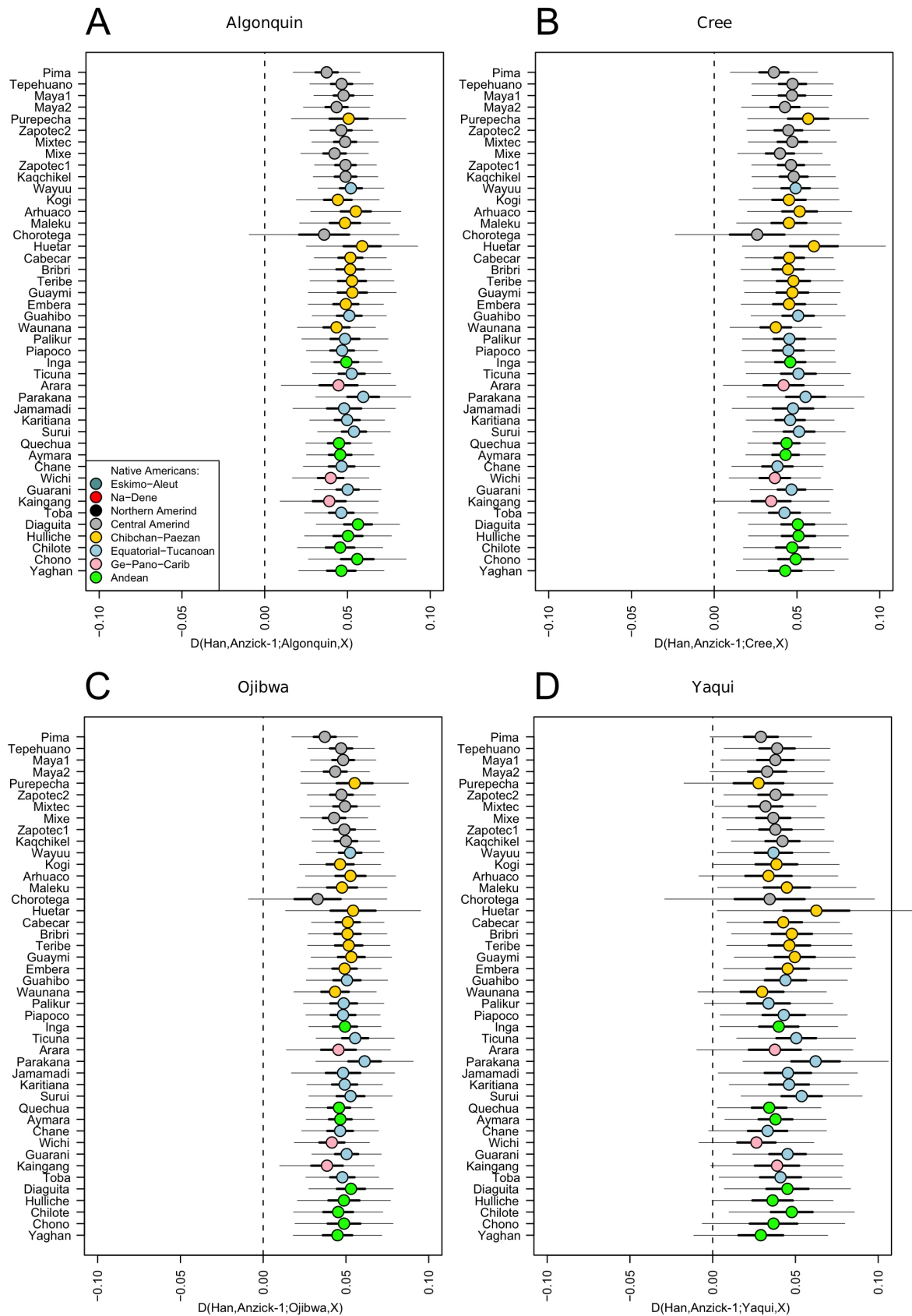
within the Y-chromosome phylogeny. Anzick-1 (circled) represents Y-chromosome haplogroup Q-L54* (xM3) (blue), which is offset by haplogroup Q-M3 (dark blue). The lineage carried by the ancient Saqqaq Palaeo-Eskimo (light blue) constitutes an outgroup to Q-L54. Each branch is labelled by an index and the number of transversion SNPs assigned to the branch (in brackets). Terminal taxa (individuals) are also labelled by population, ID and haplogroup. Branches 21 and 25 represent the most recent shared ancestry between Anzick-1 and other members of the sample. Branch 19 is considerably shorter than neighbouring branches, which have had an additional ~12,600 years to accumulate mutations.

Anzick-1



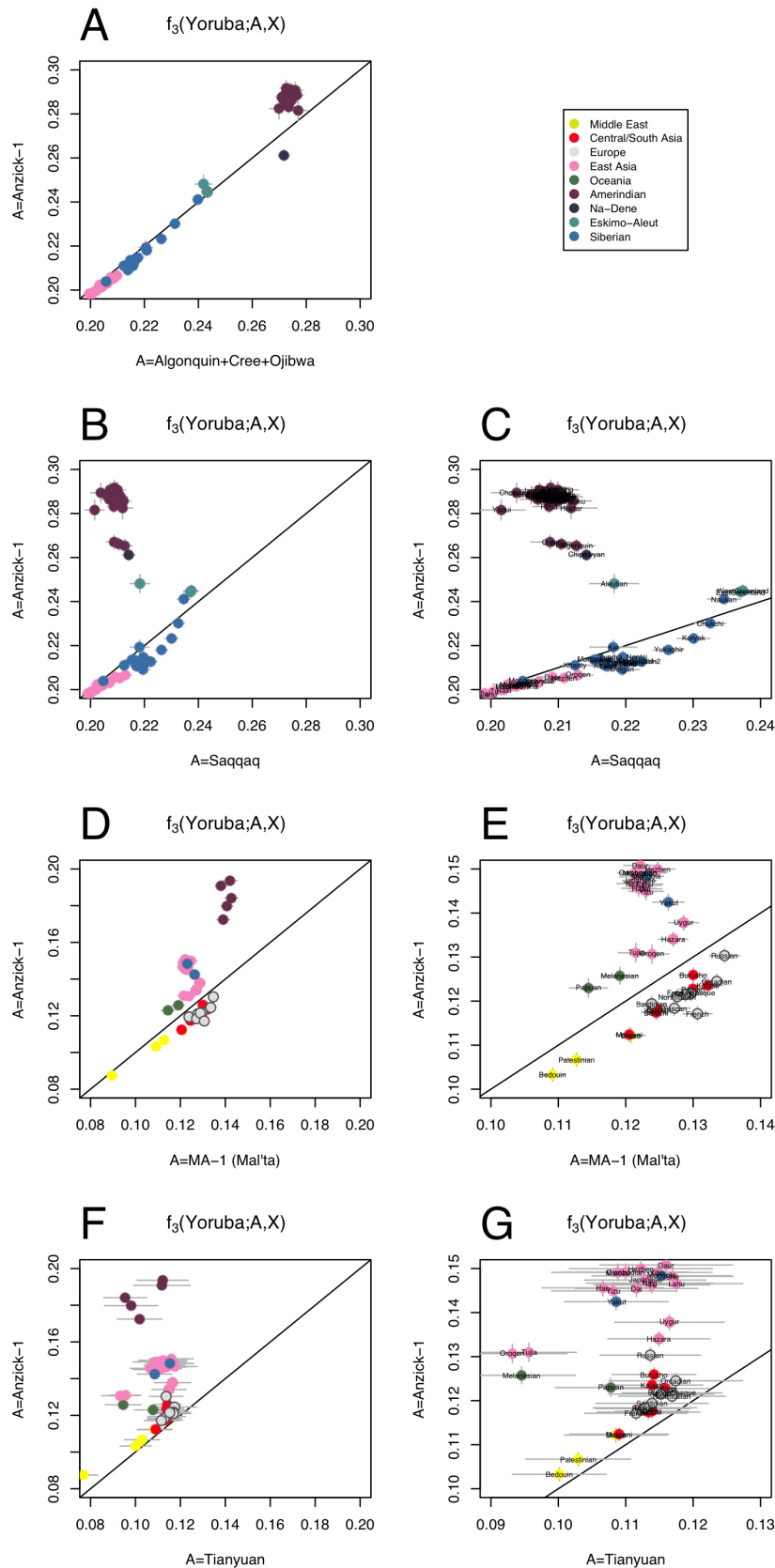
Extended Data Figure 3 | Ancestry proportions of Anzick-1 as determined by Admixture assuming the number of hypothetical ‘ancestral’ populations or genetic components, K, is 3 to 5 and 9 to 11 for a set of 135 extant Eurasian, Oceanian and New World populations. Shown are results from one of the converged runs at each K. We note that the model at K = 11 was found to have the best predictive accuracy as determined by the lowest cross

validation index values (see Supplementary Information). At each K each sample is represented by a stacked vertical bar whereas these of Anzick-1 are magnified and presented horizontally at the top. Note that irrespective of the number of genetic components, K, assumed, the Anzick-1 sample shares all the components present in different contemporary Native American populations.



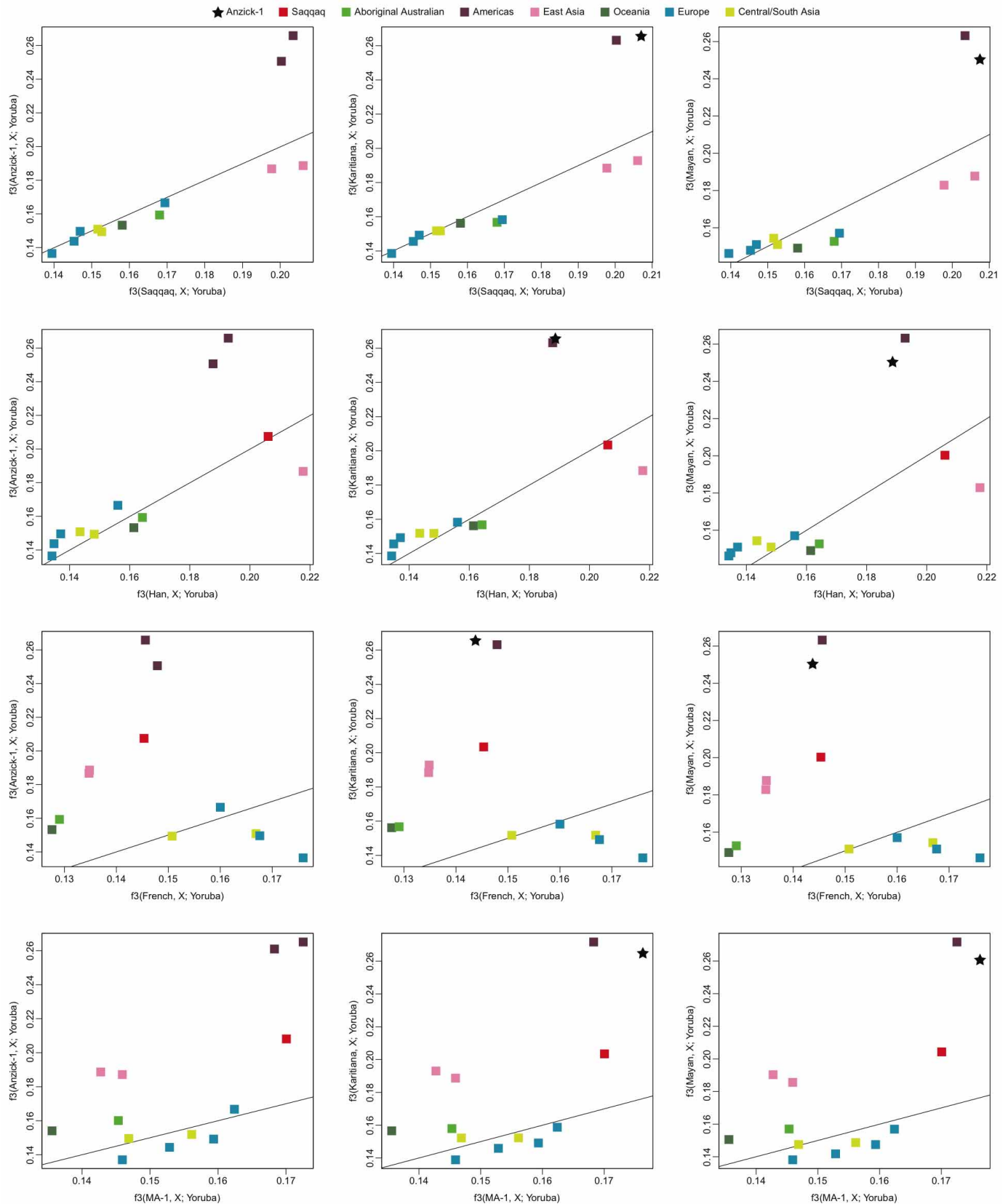
Extended Data Figure 4 | Anzick-1 is closer to Central/Southern Native Americans than Northern Native Americans. a–d, The closer relationship between Anzick-1 and Southern Native Americans compared to Algonquin, Cree, Ojibwa and a Yaqui individual is consistent for different 44 Southern and

Central Native American populations to Anzick-1. We used the test $D(\text{Han}, \text{Anzick-1}; \text{Algonquin/Cree/Ojibwa/Yaqui}, \text{Central/Southern Native Americans})$. Thick and thin whiskers represent 1 and 3 standard errors, respectively.



Extended Data Figure 5 | Outgroup f_3 -statistics contrasted for different combinations of populations. a, Shared genetic history with Anzick-1 compared to shared genetic history with the three Northern Amerind-speaking populations. b, c, Shared genetic history with the Anzick-1 individual

compared to the ~4,000-year-old Saqqaq from Greenland. d, e, Anzick-1 compared to shared genetic history with the 24,000-year-old MA-1 individual from Central Siberia. f, g, Shared genetic history with Anzick-1 compared to shared genetic history with the 40,000-year-old Tianyuan individual from China.



Extended Data Figure 6 | Pairwise outgroup f_3 statistics computed using Saqqaq, Han, French or ancient MA-1 (Mal'ta) on the x-axis and Anzick-1, Karitiana or Mayan on the y-axis. The black line indicates the $y = x$ line.

Extended Data Table 1 | Accelerator Mass Spectrometry (AMS) ^{14}C and stable isotope analyses of human bone and antler artefacts from the Anzick site, Montana

AMS LAB NO.	Chemical Fraction Dated	^{14}C Age, RC yr. \pm SD	^{14}C Date, CAL BP (95.4% C.I.)	$\delta^{15}\text{N}$ ‰ (AIR)	$\delta^{13}\text{C}$ ‰ (VPDB)	Collagen Yield (wt%)	%N	%C	C/N (atomic ratio)	Year Measured	Ref
Red, Hematite-Stained Petrous Bone (ANZICK-1)											
AA-313A	Decalcified Collagen	8690 \pm 310								1984	Stafford, 1987
AA-313B	Untreated Gelatin	10,500 \pm 400								1984	Stafford, 1987
AA-2978	Aspartic acid	10,240 \pm 120								1990	Stafford, 1991
AA-2982	Alanine	10,370 \pm 130								1990	Stafford, 1991
AA-2980	Hydroxyproline	10,710 \pm 100								1990	Stafford, 1991
AA-2979	Glutamic acid	10,820 \pm 100								1990	Stafford, 1991
AA-2981	Glycine	10,940 \pm 90								1990	Stafford, 1991
CAMS-80535	Decalcified Collagen	10,580 \pm 35		+9.1	-17.6	-	-	-	-	2001	This study
CAMS-80536	KOH-Collagen	10,525 \pm 35		+8.6	-18.1	-	-	-	-	2001	This study
CAMS-80537	Gelatin	10,610 \pm 30		+10.1	-18.0	-	-	-	-	2001	This study
CAMS-80538	XAD-Gelatin	10,705 \pm 35	12,722-12,590	-	-	-	-	-	-	2001	This study
Beta-163833	Alkali Collagen	10,780 \pm 40	12,746-12,660	-	-21.2	-	-	-	-	2006	Morrow & Gnecco, 2006
SR-8149	KOH-Collagen	-		+10.5	-18.0	22.7	15.7%	43.9%	3.26	2013	This study
SR-8149	Gelatin	-		+11.1	-17.9	22.7	14.4%	40.0%	3.24	2013	This study
SR-8151	KOH-Collagen	-		+11.0	-17.9	18.7	14.9%	41.9%	3.29	2013	This study
SR-8151	Gelatin	-		+11.6	-17.8	18.7	14.7%	41.2%	3.26	2013	This study
Antler Rods											
Specimen #118/119 UCIAMS-61661	XAD-Collagen	11,025 \pm 30	13,010-12,777	-	-	-	-	-	-	2006	This study
Beta-163832	Alkali Collagen	11,040 \pm 60	13,053-12,753	-	-22.5	-	-	-	-	2006	Morrow & Gnecco, 2006
Specimen #122 Beta-168967	Alkali Collagen	11,040 \pm 40	13,040-12,780	-	-18.2	-	-	-	-	2006	Morrow & Gnecco, 2006
	Average, 3-rods	11,035 \pm 45	13,039-12,763								

The previously published ^{14}C measurements are provided for reference. Some of the analyses were carried out to collect stable isotope and geochemistry data, and thus show no age information. Calibrated ages using IntCal 09 are given only for ^{14}C dates completed since 2006 and for chemical fractions considered the highest purity and having the greatest accuracy; calibration of chemically impure chemical fractions is not warranted. AA, NSF-Arizona Accelerator Mass Spectrometry Facility, Tucson; Beta, Beta Analytic, Florida; CAMS, Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratories, Livermore, California; SR, Stafford Research, Boulder, Colorado. '-' indicates not tested.

Extended Data Table 2 | Statistics of the genome sequence data used in the study and high-confidence genotype calls

Sample	Region	Population	Reference	Gender	Avg. Depth	Cov >1X	HomRef	Het	HomAlt	Accessible
DNK02	Africa	Dinka	Meyer <i>et al</i> , 2012	Male	24.3	0.90	1,356,439,142	1,194,183	653,791	1,358,287,116
HGDP00456	Africa	Mbuti	Meyer <i>et al</i> , 2012	Male	20.3	0.90	1,333,191,748	1,198,421	741,996	1,335,132,165
HGDP00521	Europe	French	Meyer <i>et al</i> , 2012	Male	22.6	0.90	1,352,074,026	928,373	611,106	1,353,613,505
HGDP00542	SE Asia	Papuan	Meyer <i>et al</i> , 2012	Male	21.6	0.90	1,343,127,185	745,786	753,954	1,344,626,925
HGDP00665	Europe	Sardinian	Meyer <i>et al</i> , 2012	Male	19.9	0.90	1,334,884,400	908,557	606,234	1,336,399,191
HGDP00778	Han	Han	Meyer <i>et al</i> , 2012	Male	22.3	0.90	1,349,104,846	879,299	668,621	1,350,652,766
HGDP00927	Africa	Yoruba	Meyer <i>et al</i> , 2012	Male	26.7	0.90	1,361,142,931	1,221,258	664,333	1,363,028,522
HGDP00998	S. America	Karitiana	Meyer <i>et al</i> , 2012	Male	21.3	0.90	1,343,817,819	667,456	753,001	1,345,238,276
HGDP01029	Africa	San	Meyer <i>et al</i> , 2012	Male	26.9	0.90	1,362,464,890	1,273,369	774,869	1,364,513,128
HGDP01284	Africa	Mandenka	Meyer <i>et al</i> , 2012	Male	20.6	0.90	1,338,712,364	1,210,852	646,656	1,340,569,872
HGDP01307	SE Asia	Dai	Meyer <i>et al</i> , 2012	Male	23.8	0.90	1,350,414,031	879,716	660,662	1,351,954,409
HGDP00877	S. America	Mayan	This study	Male	13.2	0.90	1,130,447,805	673,358	547,730	1,131,668,893
BI16	S. America	Karitiana	This study	Female	23.4	0.89	1,278,750,988	664,223	684,479	1,280,099,690
Avar	SC Asia	Avar	Raghavan <i>et al</i> , 2013	Male	12.8	0.83	622,360,039	468,944	299,356	623,128,339
Indian	SC Asia	Indian	Raghavan <i>et al</i> , 2013	Female	15.9	0.89	861,422,992	643,238	389,937	862,456,167
Mari	SC Asia	Mari	Raghavan <i>et al</i> , 2013	Male	12.0	0.87	235,202,480	199,301	120,287	235,522,068
Tadjik	SC Asia	Tajik	Raghavan <i>et al</i> , 2013	Male	16.4	0.88	939,846,482	687,749	431,934	940,966,165
AusAboriginal	SE Asia	Australian Aboriginal	Rasmussen <i>et al</i> , 2011	Male	6.3	0.50	356,126,648	240,732	220,901	356,588,281
Saqqaq	Greenlander	Saqqaq	Rasmussen <i>et al</i> , 2010	Male	12.8	0.79	515,972,797	307,585	306,065	516,586,447
Anzick-1	N. America	Clovis	This study	Male	14.4	0.87	780,953,901	864,870	409,731	782,228,502

AvgDepth, average depth; Cov >1, fraction of the genome covered with at least 1 read; Het: heterozygote call; HomAlt, Homozygote alternative call; HomRef, Homozygote reference call.

Extended Data Table 3 | Sequencing and mapping statistics for the Anzick-1 libraries

Library	Raw Reads	Per cent human reads	Mapped human reads \geq Q30	Final BAM	Per cent clonality
C10	55,495,972	0.5%	289,526	96,136	67%
C121121_36	2,000,862,136	24.6%	491,779,235	116,018,652	76%
C121121_37	1,938,412,150	24.8%	481,682,594	100,596,004	79%
C28	412,441,694	27.8%	114,736,785	89,356,892	22%
C30_8	943,369,850	22.7%	214,367,643	21,016,810	90%
C30_8_26	1,234,827,852	23.5%	289,965,256	18,404,080	94%
C30_8_48	1,116,970,391	23.4%	261,200,675	15,489,743	94%
C7	531,754,382	12.0%	63,602,547	8,025,652	87%
C8	298,045,766	10.5%	31,415,268	7,273,279	77%
C9	247,005,820	0.9%	2,234,390	171,718	92%
CL16	614,007,221	22.1%	135,844,895	28,405,912	79%
CL18	803,188,531	21.6%	173,273,845	35,132,866	80%
CL20	695,954,288	21.7%	150,806,716	29,319,798	81%
CL32	477,347,711	28.2%	134,386,357	108,903,306	19%
CL33	791,303,957	21.4%	169,518,285	25,024,524	85%
Cpro_42	201,998,907	8.3%	16,846,370	3,381,783	80%
Cpro_43	305,882,477	8.3%	25,418,416	2,896,283	89%
Crib_34	323,616,468	0.9%	2,983,594	594,741	80%
Crib_39	281,843,911	0.9%	2,447,883	302,525	88%
Total	13,274,329,484	20.8%	2,762,800,280	610,410,704	78%

Extended Data Table 4 | Site patterns and drift estimates for non-African populations versus the Anzick-1 sample

Including all sites								
ID	Population	n(AAAa)	n(AaAA)	n(AAaa)	n(AaAa)	n(AAAA)	Drift in population	Drift in Anzick-1
HGDP00877 (masked)	Mayan	123095	112507	38900	76913	1064893	0.0294	0.0000
HGDP00998	Karitiana	222029	162274	77002	111472	1791319	0.1581	0.0000
BI16	Karitiana	196546	150372	66032	105219	1630510	0.1258	0.0000
Saqqaq	Saqqaq	152815	129619	71706	71545	1179386	0.1911	0.0979
HGDP00778	Han	210095	237386	97103	124521	1703214	0.0553	0.1266
HGDP01307	Dai	211459	239636	98327	125610	1713361	0.0554	0.1283
HGDP00542	Papuan	237801	200257	140039	95850	1688750	0.2725	0.1802
Aboriginal	Aboriginal	116797	100838	67763	48350	847186	0.2569	0.1774
HGDP00521	French	212922	265527	112889	122188	1661041	0.0617	0.1890
HGDP00665	Sardinian	211703	261689	116240	118827	1633910	0.0772	0.1984
Avar	Avar	161692	197302	84076	92078	1262525	0.0653	0.1804
Mari	Mari	67063	81810	32757	40631	547086	0.0397	0.1554
Tadjik	Tadjik	195832	247682	99654	118142	1567438	0.0385	0.1748
Indian	Indian	170384	214686	86426	103309	1371359	0.0382	0.1722
Excluding sites where a C and T or G and A were observed								
ID	Population	n(AAAa)	n(AaAA)	n(AAaa)	n(AaAa)	n(AAAA)	Drift in population	Drift in Anzick-1
HGDP00877 (masked)	Mayan	38605	35517	12138	24518	334804	0.0232	0.0000
HGDP00998	Karitiana	68442	51065	24000	34758	554162	0.1519	0.0000
BI16	Karitiana	60878	47002	20607	33087	509629	0.1216	0.0000
Saqqaq	Saqqaq	46180	39448	21528	21683	358284	0.1869	0.0974
HGDP00778	Han	64920	74011	29884	38635	527612	0.0507	0.1273
HGDP01307	Dai	65190	74986	30472	39117	530308	0.0501	0.1316
HGDP00542	Papuan	73369	62598	43620	29859	522442	0.2693	0.1842
Aboriginal	Aboriginal	34529	30008	20184	14512	250288	0.2544	0.1787
HGDP00521	French	65748	83094	34912	38023	514107	0.0563	0.1915
HGDP00665	Sardinian	65549	81338	35819	36721	506155	0.0749	0.1987
Avar	Avar	49028	60027	25643	27920	384344	0.0657	0.1825
Mari	Mari	19666	24153	9541	12091	160779	0.0334	0.1531
Tadjik	Tadjik	60041	76457	30652	36532	483570	0.0357	0.1758
Indian	Indian	52265	65827	26564	31632	420851	0.0391	0.1728

Top, data based on conditioning on African polymorphism. Bottom, data based on conditioning on African polymorphism and removing sites where a C and T or a G and A were observed.

Intranasal epidermal growth factor treatment rescues neonatal brain injury

Joseph Scafidi^{1,2}, Timothy R. Hammond^{1,3}, Susanna Scafidi⁴, Jonathan Ritter¹, Beata Jablonska¹, Maria Roncal¹, Klara Szigeti-Buck⁵, Daniel Coman⁶, Yuegao Huang⁶, Robert J. McCarter Jr⁷, Fahmeed Hyder⁶, Tamas L. Horvath⁵ & Vittorio Gallo¹

There are no clinically relevant treatments available that improve function in the growing population of very preterm infants (less than 32 weeks' gestation) with neonatal brain injury. Diffuse white matter injury (DWMI) is a common finding in these children and results in chronic neurodevelopmental impairments^{1,2}. As shown recently, failure in oligodendrocyte progenitor cell maturation contributes to DWMI³. We demonstrated previously that the epidermal growth factor receptor (EGFR) has an important role in oligodendrocyte development⁴. Here we examine whether enhanced EGFR signalling stimulates the endogenous response of EGFR-expressing progenitor cells during a critical period after brain injury, and promotes cellular and behavioural recovery in the developing brain. Using an established mouse model of very preterm brain injury, we demonstrate that selective overexpression of human EGFR in oligodendrocyte lineage cells or the administration of intranasal heparin-binding EGF immediately after injury decreases oligodendroglia death, enhances generation of new oligodendrocytes from progenitor cells and promotes functional recovery. Furthermore, these interventions diminish ultrastructural abnormalities and alleviate behavioural deficits on white-matter-specific paradigms. Inhibition of EGFR signalling with a molecularly targeted agent used for cancer therapy demonstrates that EGFR activation is an important contributor to oligodendrocyte regeneration and functional recovery after DWMI. Thus, our study provides direct evidence that targeting EGFR in oligodendrocyte progenitor cells at a specific time after injury is clinically feasible and potentially applicable to the treatment of premature children with white matter injury.

Chronic neonatal hypoxia is a clinically relevant model of premature brain injury caused by insufficient gas exchange from poor lung development⁵. This 'hypoxic' state is a major contributor to DWMI—a common finding in infants born very preterm (VPT)—resulting in sensorimotor deficits that persist throughout their lifetime^{1,2,6}. We used a mouse model of chronic hypoxia, which replicates DWMI and other neuropathological hallmarks of brain injury resulting from premature birth^{7–9}.

The cellular and molecular mechanisms underlying DWMI in VPT children—and in chronic hypoxia—are unknown. It has been previously demonstrated that enhanced EGFR signalling in white matter oligodendrocyte lineage cells promotes their proliferation, migration, myelination and remyelination in the adult^{4,10}. We observed a significant increase in endogenous EGF levels in the white matter after chronic hypoxia (Extended Data Fig. 1). Therefore, we compared oligodendrocyte development in white matter injury and recovery in 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNP) enhanced green fluorescent protein (eGFP) mice (Rep mice) and Rep mice in which human (h)EGFR was overexpressed in the oligodendrocyte lineage under the CNP promoter (Rep-hEGFR mice)^{4,11–13}. Hypoxia decreased myelin basic protein (MBP) expression in the white matter of Rep mice, but not in Rep-hEGFR mice (Fig. 1a–e). At postnatal day (P)60, MBP expression

recovered in the hypoxia-treated Rep group (Fig. 1e). At P11, chronic hypoxia did not cause any change in the number of Rep⁺Olig2⁺ cells and mature (Rep⁺CC1⁺) oligodendrocytes (Fig. 1f). At P18, we observed a decrease in Rep⁺Olig2⁺ and Rep⁺CC1⁺ oligodendrocytes in the white matter of hypoxia-treated Rep mice (Fig. 1g), but no change in the Rep-hEGFR mice. Oligodendrocyte recovery was evident by P60 in the hypoxia-treated Rep group (Fig. 1h).

There was an increase in apoptosis of oligodendrocyte lineage cells in Rep mice after hypoxia treatment at P11 and P18, but no change at P60 (Extended Data Fig. 2e). No significant apoptosis was observed in Rep-hEGFR mice (Extended Data Fig. 2e). Hypoxia caused an increase in the number of propidium iodide (PI)⁺ cells (data not shown) and Rep⁺PI⁺ cells (Extended Data Fig. 2a–d, f). This increase was not observed in Rep-hEGFR mice. These results indicate that enhanced EGFR expression prevents oligodendrocyte loss by decreasing cell death after exposure to chronic hypoxia.

We next assessed the effects of hypoxia on oligodendrocyte progenitor (Rep⁺NG2⁺) cells (OPCs) in the white matter (Extended Data Fig. 2g–k). Enhanced hEGFR expression caused an increase in Rep⁺NG2⁺ OPCs at P11 and P18 (normoxia Rep mice versus normoxia Rep-hEGFR mice; Extended Data Fig. 2k). Hypoxia caused a significant increase in white matter OPCs in both Rep and Rep-hEGFR mice at the same ages (Extended Data Fig. 2k). Similar findings were obtained after assessing proliferation of Rep⁺ oligodendrocyte lineage cells (Extended Data Fig. 2l). Enhanced hEGFR expression increased Rep⁺NG2⁺ OPC proliferation in the normoxia group, and had an additive effect on hypoxia-induced OPC proliferation (Extended Data Fig. 2m). Enhanced hEGFR expression increased oligodendrogenesis at P18, but, at P30, no difference was evident between hypoxia-treated Rep and hypoxia-treated Rep-hEGFR mice (Fig. 1i). These results indicate that enhanced hEGFR expression in oligodendrocytes promoted the generation of new oligodendrocytes after chronic hypoxia.

We used electron microscopy to determine whether chronic hypoxia caused myelination abnormalities, and to assess whether EGFR overexpression rescued these abnormalities. (Fig. 1j–o). At P60, when oligodendrocyte cell numbers and MBP expression recovered, myelination was still abnormal after hypoxia (Fig. 1j–o). Hypoxia caused a significant increase in the g ratio and hEGFR expression prevented this increase (Fig. 1n, o).

Next, we investigated behavioural deficits resulting from DWMI after perinatal chronic hypoxia by using subcortical white-matter-dependent sensorimotor behavioural tests (complex wheel and inclined beam-walking task)^{14–18}. In the complex wheel task, there was no difference in training maximum velocity (V_{\max}) between all four groups (Fig. 1p). On day 15, all four groups had a decline in V_{\max} ; however, the largest decline was in the hypoxia-treated Rep group (Fig. 1p). The hypoxia-treated Rep group performed poorly on the complex wheel (days 15–21),

¹Center for Neuroscience Research, Children's National Medical Center, Washington DC 20010, USA. ²Department of Neurology, Children's National Medical Center, Washington DC 20010, USA. ³Institute for Biomedical Sciences, The George Washington University, Washington DC 20052, USA. ⁴Department of Anesthesiology & Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. ⁵Department of Neurobiology, Yale University, New Haven, Connecticut 06520, USA. ⁶MRRC, Department of Diagnostic Radiology, Yale University, New Haven, Connecticut 06520, USA. ⁷Center for Translational Science, Children's National Medical Center, Washington DC 20010, USA.

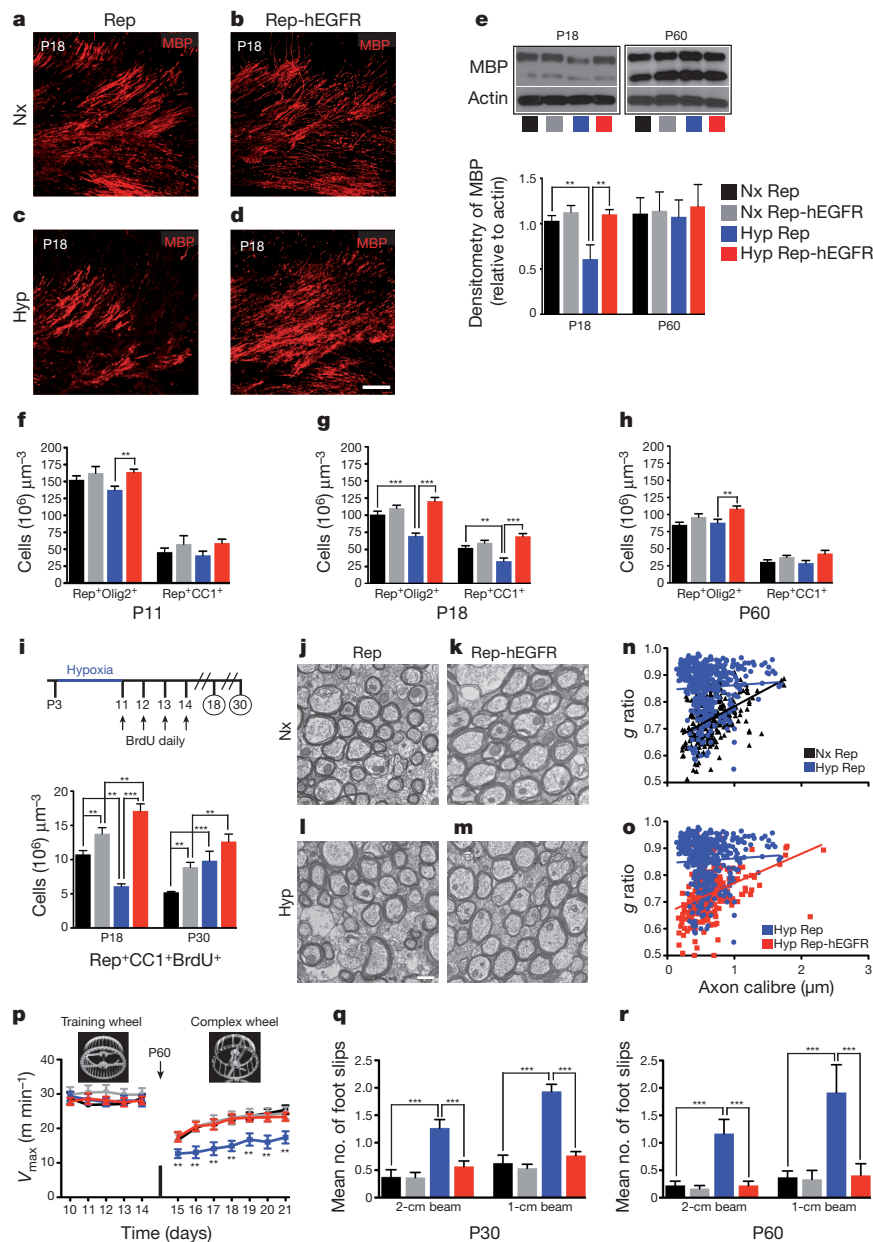


Figure 1 | Enhanced EGFR expression in oligodendrocyte lineage cells prevents oligodendrocyte and myelin loss, and ultrastructural and behavioural deficits caused by neonatal hypoxia. **a–d**, Confocal images of white matter immunostained for MBP. Hyp, hypoxia-treated group; Nx, normoxia group. Scale bar, 50 μm. **e**, Western blot of white matter tissue ($n = 4$ mice per group and per age except P18 normoxia and hypoxia-exposed mice, $n = 5$; one-way analysis of variance (ANOVA), Bonferroni post-hoc test for individual comparisons). **f–h**, Number of Rep⁺Olig2⁺ and Rep⁺CC1⁺ cells. **i**, Number of newly generated oligodendrocytes in white matter. **f–i**, $n = 4$ mice per group and per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons. **j–m**, Electron microscopy images from P60 white matter. Scale bar, 0.5 μm. **n, o**, Scatter plots of g ratios of individual axons relative to axon diameters ($n = 3$ mice per group; one-way ANOVA of all four groups with post-hoc unpaired t -tests). **p**, V_{max} (m min⁻¹) over time (days) on the complex wheel (linear regression comparison of slopes between all four groups; post-hoc comparison of individual days between groups: Nx Rep, $n = 12$; Nx Rep-hEGFR, $n = 9$; Hyp Rep, $n = 12$; Hyp Rep-hEGFR, $n = 10$). **q**, Naive mice were tested on the 2-cm- and 1-cm-width inclined beam-walking task (Poisson multiple regression analysis; Nx Rep, $n = 8$; Nx Rep-hEGFR, $n = 10$; Hyp Rep, $n = 10$; Hyp Rep-hEGFR, $n = 10$). **r**, A separate group of naive mice was assessed (Poisson multiple regression analysis; Nx Rep, $n = 9$; Nx Rep-hEGFR, $n = 8$; Hyp Rep, $n = 10$; Hyp Rep-hEGFR, $n = 9$). Line graph and histograms are presented as means \pm standard error of the mean (s.e.m.). * $P = 0.05$; ** $P < 0.05$; *** $P < 0.01$.

as compared to the other three groups (Fig. 1p), suggesting altered subcortical white matter integrity.

The inclined beam-walking task^{17,18} requires subcortical white matter integrity (Extended Data Fig. 3a–c) and no difference was observed between Rep mice and C57BL/6 mice (Extended Data Fig. 3d). At P30, the hypoxia-treated Rep group displayed more foot slips than the normoxia Rep group (Fig. 1q). Hypoxia-treated Rep-hEGFR mice showed no significant increase in the number of foot slips as compared with the normoxia Rep-hEGFR group. At P60, the hypoxia-treated Rep group continued to show an increase in number of foot slips, and no difference was seen in the hypoxia-treated Rep-hEGFR group (Fig. 1r). These behavioural studies confirm that exposure to chronic hypoxia during a critical period in myelin development results in poor performance on white-matter-specific behavioural tasks. Enhanced EGFR activity prevents the effects of hypoxia, strongly suggesting that EGFR signalling in oligodendrocyte lineage cells has a crucial role in white matter recovery after perinatal injury.

We directly tested the role of endogenous EGFR signalling in oligodendrocyte recovery after hypoxia by gefitinib administration from P12–P18 (Fig. 2a). Gefitinib, a specific EGFR antagonist¹⁹, caused a small

reduction in basal phosphorylated (p)EGFR levels in the white matter of normoxia mice at P18, but completely prevented the increase in pEGFR induced by hypoxia (Fig. 2b). Gefitinib caused a decrease of white matter Rep⁺Olig2⁺ and Rep⁺CC1⁺ oligodendrocyte lineage cells in normoxia mice (Fig. 2c). Hypoxia resulted in a decrease of these cell populations, which were further reduced by gefitinib (Fig. 2c). This decrease was attributed to a significant increase in oligodendrocyte apoptosis in both normoxia and hypoxia groups (Fig. 2d). Gefitinib also decreased NG2⁺ OPCs in the white matter of normoxia mice, and prevented the increase observed after hypoxia (Fig. 2e). Similarly, gefitinib decreased white matter oligodendrocyte lineage cell proliferation in normoxia mice and prevented the proliferative response observed after hypoxia (Fig. 2f). Gefitinib also decreased Sox2- and Ascl1-expressing progenitors (Extended Data Fig. 4a–j). At P30, gefitinib still caused a decrease in oligodendrocyte lineage cells and mature oligodendrocytes in normoxia mice, and prevented oligodendrocyte recovery and oligodendrogenesis observed after hypoxia (Fig. 2g, h). Finally, gefitinib prevented the recovery in CNPase and MBP protein levels observed after hypoxia at P30 (Fig. 2i). These results confirm that endogenous EGFR signalling is important in white matter cellular and biochemical recovery after hypoxia.

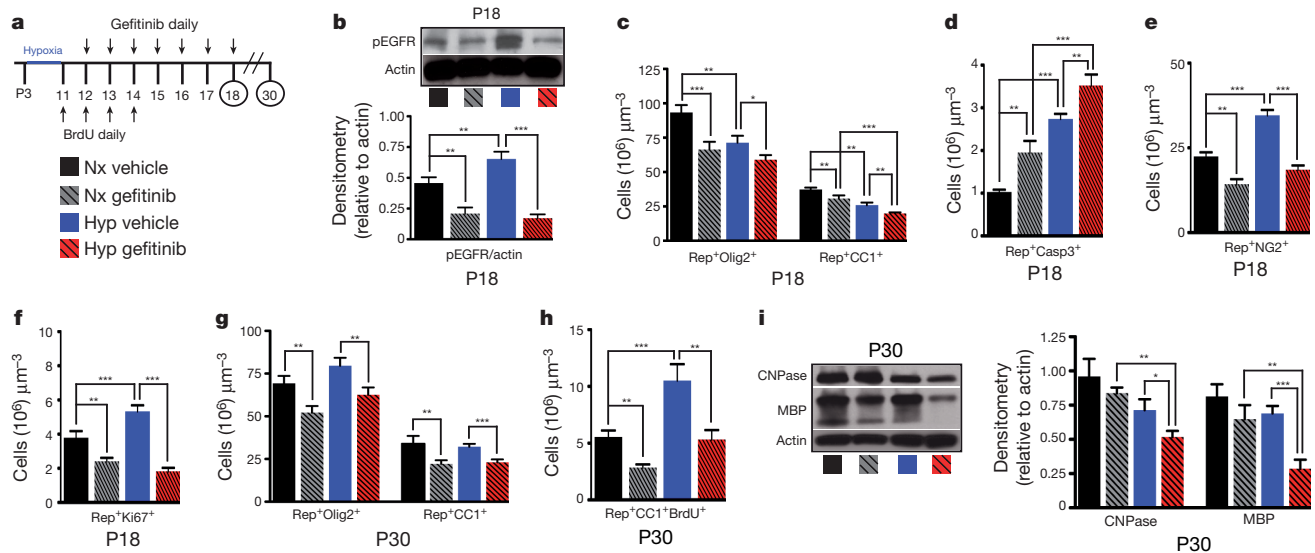
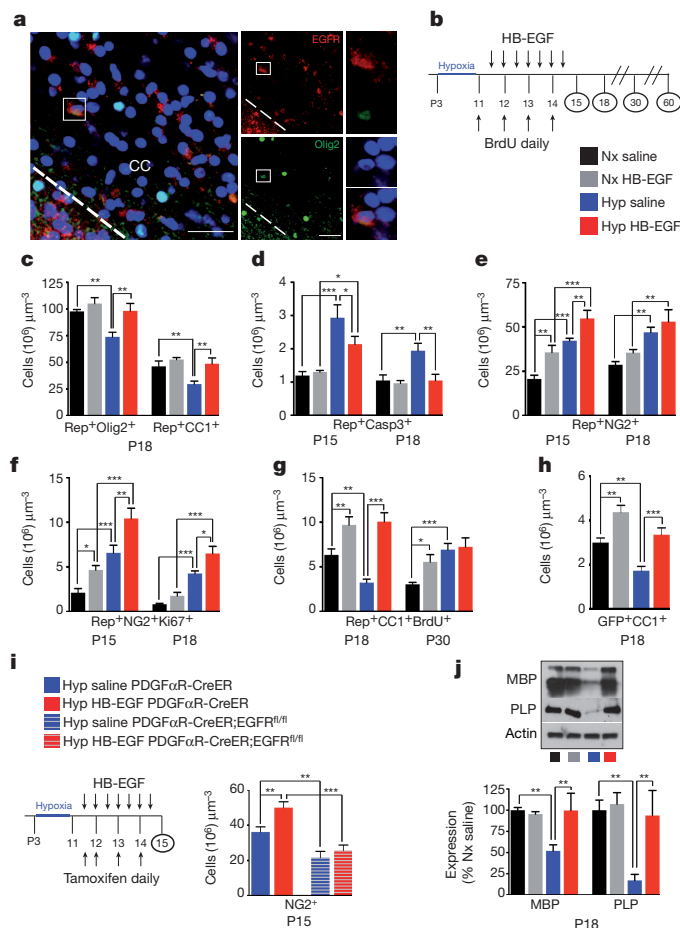


Figure 2 | EGFR activity is crucial for white matter recovery after neonatal hypoxia. **a**, Protocol of gefitinib and BrdU administration. Different groups are indicated. Hyp, hypoxia treated; Nx, normoxia. **b**, Western blot of white matter shows that gefitinib decreased pEGFR in the normoxia group and prevented the increase in pEGFR after hypoxia ($n = 5$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **c**, Counts of Rep⁺Olig2⁺ and Rep⁺CC1⁺ cells in white matter. **d**, Gefitinib increased cell apoptosis in Nx and Hyp groups. **e**, f, Gefitinib decreased Rep⁺NG2⁺ OPCs (e) and oligodendrocyte-lineage cell proliferation (f) in normoxia mice and prevented

hypoxia-induced increase in OPC and oligodendrocyte-lineage cell proliferation. **g**, Long-term effects of gefitinib on Rep⁺Olig2⁺ and Rep⁺CC1⁺ cells. **h**, Gefitinib decreased newly generated Rep⁺CC1⁺ oligodendrocytes in normoxia mice and prevented oligodendrogenesis after hypoxia. **i**, Gefitinib prevented recovery of CNPase and MBP expression after hypoxia. **c–i**, $n = 4$ mice per all groups and per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons. All histograms are presented as means \pm s.e.m. * $P = 0.05$; ** $P < 0.05$; *** $P < 0.01$.



We wanted to examine whether directly targeting EGFR with a selective ligand (recombinant heparin-binding EGF (HB-EGF)) delivered through the intranasal route promoted cellular recovery of white matter oligodendrocytes after hypoxia. The clinical relevance of targeting endogenous oligodendrocytes was demonstrated by the presence of EGFR-expressing oligodendrocyte lineage cells (Olig2⁺EGFR⁺) in neonatal preterm human white matter (Fig. 3a). The intranasal route allows rapid drug delivery directly to the brain from the nasal mucosa^{20–22}. HB-EGF via the intranasal route enters the brain^{20,23} and acts on the white matter (Extended Data Fig. 5a–c), where Rep⁺pEGFR⁺ cells could be identified (Extended Data Fig. 5d). FACS purification of hypoxia-exposed white matter Rep⁺ cells after HB-EGF treatment directly demonstrated activation of EGFR in oligodendrocyte lineage cells (Extended Data Fig. 5e). Treatment with seven doses of intranasal HB-EGF from P11–P14 (Fig. 3b)

Figure 3 | Intranasal HB-EGF accelerates oligodendrocyte regeneration and promotes cellular recovery in white matter after neonatal hypoxia. **a**, Confocal image of the corpus callosum (CC) in a preterm infant, highlighting an Olig2⁺EGFR⁺ cell (box). Scale bars, 50 μ m. **b**, Protocol of intranasal HB-EGF/BrdU administration and tissue collection. Different groups are indicated. Hyp, hypoxia treated; Nx, normoxia. **c**, Number of white matter Rep⁺Olig2⁺ and Rep⁺CC1⁺ cells in HB-EGF-treated mice. **d**, HB-EGF attenuated or prevented the effects of hypoxia on oligodendrocyte apoptosis. **e**, HB-EGF had an additive effect on hypoxia-induced increase of Rep⁺NG2⁺ OPCs at P15, but not at P18. **f**, HB-EGF had an additive effect on hypoxia-induced increase of OPC proliferation. **g**, HB-EGF promoted oligodendrogenesis after hypoxia at P18. **h**, Fate mapping of OPCs (PDGF α R-CreER;Z/EG (GFP) mouse) demonstrated that oligodendrogenesis occurred from PDGF α R⁺ cells. **c–h**, $n = 4$ mice per group and per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons. **i**, Removal of EGFR in PDGF α R-expressing OPCs (PDGF α R-CreER;EGFR^{fl/fl};Z/EG mouse) caused a decrease in NG2⁺ OPCs after hypoxia and prevented the effects of HB-EGF ($n = 4$ mice per group except hypoxia HB-EGF PDGF α R-CreER;EGFR^{fl/fl} group, $n = 3$; one-way ANOVA of all four groups with post-hoc unpaired t -tests). **j**, HB-EGF promoted recovery in white matter MBP and PLP protein levels after hypoxia (western blot; $n = 6$ mice per group except Nx HB-EGF, $n = 4$; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). All histograms are presented as means \pm s.e.m. * $P = 0.05$; ** $P < 0.05$; *** $P < 0.01$.

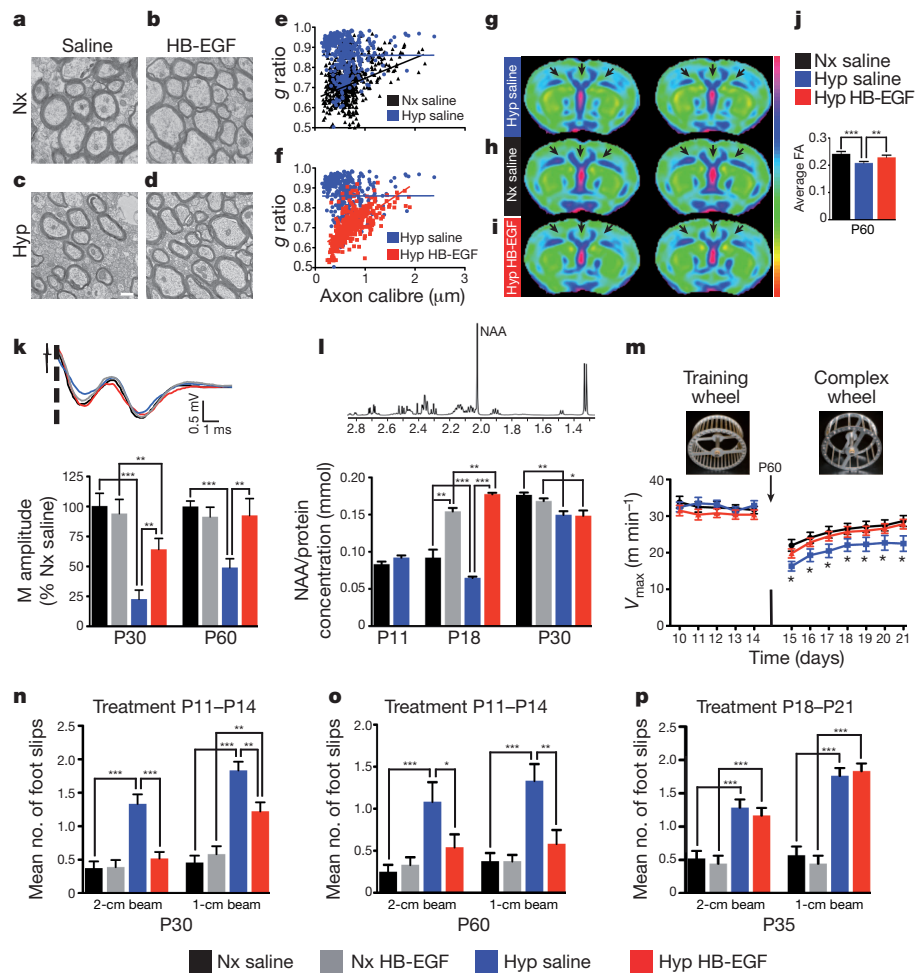


Figure 4 | Intranasal HB-EGF promotes ultrastructural, physiological and behavioural recovery in white matter after neonatal hypoxia. **a–d**, Electron microscopy images of P30 white matter (a–d) and scatter plots of *g* ratios of individual axons relative to axon diameters (e, f) ($n = 3$ mice per group; one-way ANOVA of all four groups with post-hoc unpaired *t*-tests). Hyp, hypoxia treated; Nx, normoxia. Scale bar, 0.5 μm . **g–j**, *Ex vivo* DTI analysis (representative averaged FA maps) shows that HB-EGF attenuates hypoxia-induced reduction of FA values in the corpus callosum, cingulum and external capsule (arrows point to regions of interest; $n = 5$ for each group). **k**, CAP extracellular recordings. Representative waveforms and histograms show that the Hyp HB-EGF group had larger myelinated (M) amplitudes compared with the Hyp saline group ($n = 7$ per group per age except P60 Nx HB-EGF, $n = 5$; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **l**, *Ex vivo* measurements of white matter NAA using ^1H -nuclear magnetic resonance (^1H -NMR) spectroscopy. Representative ^1H -NMR spectra

showing the NAA peak. At P18 there was significantly less NAA in the Hyp saline group compared with the Hyp HB-EGF group ($n = 5$ per group per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons, except P11 unpaired *t*-test). **m**, On the complex wheel, Hyp HB-EGF mice performed similarly to the Nx saline group (linear regression comparison of slopes between all three groups and post-hoc comparison of individual days between groups; Nx saline, $n = 8$; Hyp saline, $n = 12$; Hyp HB-EGF, $n = 10$). **n, o**, On the inclined beam-walking task, HB-EGF treatment (P11–P14) decreased the number of foot slips observed in hypoxia-exposed mice (Poisson multiple regression analysis; all groups were $n = 8$, except for the P30 Hyp HB-EGF group, $n = 9$). **p**, Delayed HB-EGF administration at P18–P21 resulted in no difference in foot slips (Poisson multiple regression analysis; Nx saline and Nx HB-EGF $n = 7$; Hyp saline and Hyp HB-EGF, $n = 8$). All histograms are presented as means \pm s.e.m. * $P < 0.05$; ** $P < 0.05$; *** $P < 0.01$.

prevented white matter oligodendrocyte lineage cell loss after hypoxia (Fig. 3c). Rep⁺ cells were stained for caspase 3 in all groups. At all ages examined, HB-EGF treatment reduced (P15) or prevented (P18) the effects of hypoxia on oligodendrocyte cell death (Fig. 3d).

HB-EGF treatment also caused an increase in white matter Rep⁺NG2⁺ OPCs in both normoxia and hypoxia groups, and enhanced OPC proliferation²⁴ (Fig. 3e, f). To determine whether oligodendrocyte recovery was a result of increased oligodendrogenesis from this expanded pool of proliferative OPCs, we performed BrdU-pulse chase labelling of newly generated oligodendrocytes. HB-EGF increased the number of Rep⁺CC1⁺BrdU⁺ cells under normoxic and hypoxic conditions at P18 (Fig. 3g). At P30, the additive effect of hypoxia and HB-EGF was not as evident (Fig. 3g). Genetic lineage tracing of oligodendrocytes *in vivo* by using the platelet-derived growth factor- α receptor (PDGF α R-CreER:Z/EG) (GFP⁺) reporter mouse²⁵ (Extended Data Fig. 6 and Fig. 3h) confirmed that hypoxia and HB-EGF treatment have an additive effect

on the generation of Olig2-expressing cells (Extended data Fig. 6j). Conversely, PDGF α R-driven EGFR deletion in OPCs prevented the cellular effects of HB-EGF on NG2⁺ OPCs (Fig. 3i). HB-EGF treatment after hypoxia also resulted in recovery of MBP and proteolipid protein (PLP) expression (Fig. 3j). These results indicate that, after hypoxia, HB-EGF promotes expansion of the OPC pool, and oligodendrocyte regeneration and maturation.

Electron microscopy analysis revealed that HB-EGF treatment rescued the increase in *g* ratio observed in hypoxia-treated mice and partially prevented the decrease in percentage of myelinated axons (Fig. 4a–f and Extended Data Fig. 7). Diffusion tensor imaging (DTI) demonstrated that, at P60, fractional anisotropy (FA) values are significantly decreased in the corpus callosum, cingulum and external capsule regions of hypoxia-treated mice, but not in the HB-EGF-treated hypoxia group (Fig. 4g–j). Electrophysiological analysis of extracellular compound action potentials (CAPs) demonstrated that HB-EGF prevented the decrease in

amplitude of myelinated axons observed after hypoxia (Fig. 4k). Finally, analysis of *N*-acetylaspartate (NAA) in the white matter showed decreased levels at P18 and P30, which was prevented by HB-EGF (Fig. 4l and Extended Data Fig. 8).

On the complex wheel, hypoxia-exposed HB-EGF-treated mice showed a similar performance to the normoxia saline-treated group (Fig. 4m). Furthermore, in the inclined beam-walking task, HB-EGF treatment completely prevented hypoxia-induced behavioural deficit, as tested on the 2-cm-width beam, and reduced the effects of hypoxia on performance on the 1-cm-width beam (Fig. 4n, o). Importantly, when HB-EGF treatment was performed at P18–P21, it had no effect on the hypoxia-induced behavioural phenotype (Fig. 4p and Extended Data Fig. 9).

Importantly, HB-EGF strongly inhibited hypoxia-induced upregulation of Notch signalling elements (Extended Data Fig. 10a–c) and functional activation of Notch in white matter Olig2⁺ oligodendrocyte lineage cells (Extended Data Fig. 10d–h), which could be at least in part responsible for delayed oligodendrocyte maturation after perinatal injury²⁶. These results are consistent with the notion that HB-EGF accelerated OPC maturation after hypoxia through inhibition of Notch¹¹, and indicates that HB-EGF treatment promotes functional recovery during a critical developmental time window for effective therapeutic intervention.

Our results reveal that activating EGF/EGFR signalling promotes cellular and functional recovery after neonatal brain injury. Enhancing EGFR signalling through overexpression of the EGFR prevents DWMI, promotes the generation of new oligodendrocytes and prevents behavioural deficits in different white-matter-related tasks. Furthermore, a brief pharmacological treatment that targets endogenous EGFRs using a clinically feasible (intranasal) mode of entry during a critical window promotes cellular, developmental and myelin structural improvement, and behavioural recovery. Intranasal treatment is a plausible route to introduce sufficient HB-EGF into the brain and white matter of critically ill VPT infants.

METHODS SUMMARY

Data presented are from male mice on a C57BL/6 background. Mice underwent chronic perinatal hypoxia from P3–P11, as previously described^{27,28}. In the first set of experiments, Rep mice that did not express hEGFR were used as littermate controls. The EGFR antagonist gefitinib (AstraZeneca) was administered intraperitoneally at 75 mg kg⁻¹ body weight once daily for 7 days. Normoxia and hypoxia control groups received vehicle on the same days and times. HB-EGF was administered intranasally at a concentration of 100 ng g⁻¹ body weight in 5 µl increments separated by 10 min. Equal volumes of saline were used as a vehicle control.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 October 2012; accepted 13 November 2013.

Published online 25 December 2013.

- Ment, L. R., Hirtz, D. & Huppi, P. S. Imaging biomarkers of outcome in the developing preterm brain. *Lancet Neurol.* **8**, 1042–1055 (2009).
- Woodward, L. J., Anderson, P. J., Austin, N. C., Howard, K. & Inder, T. E. Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *N. Engl. J. Med.* **355**, 685–694 (2006).
- Buser, J. R. et al. Arrested preoligodendrocyte maturation contributes to myelination failure in premature infants. *Ann. Neurol.* **71**, 93–109 (2012).
- Aguirre, A., Dupree, J. L., Mangin, J. M. & Gallo, V. A functional role for EGFR signaling in myelination and remyelination. *Nature Neurosci.* **10**, 990–1002 (2007).
- Anjari, M. et al. The association of lung disease with cerebral white matter abnormalities in preterm infants. *Pediatrics* **124**, 268–276 (2009).
- Skranes, J. et al. Clinical findings and white matter abnormalities seen on diffusion tensor imaging in adolescents with very low birth weight. *Brain* **130**, 654–666 (2007).
- Scafidi, J., Fagel, D. M., Ment, L. R. & Vaccarino, F. M. Modeling premature brain injury and recovery. *Int. J. Dev. Neurosci.* **27**, 863–871 (2009).

- Silbereis, J. C., Huang, E. J., Back, S. A. & Rowitch, D. H. Towards improved animal models of neonatal white matter injury associated with cerebral palsy. *Dis. Model. Mech.* **3**, 678–688 (2010).
- Back, S. A. et al. Late oligodendrocyte progenitors coincide with the developmental window of vulnerability for human perinatal white matter injury. *J. Neurosci.* **21**, 1302–1312 (2001).
- Ivkovic, S., Canoll, P. & Goldman, J. E. Constitutive EGFR signaling in oligodendrocyte progenitors leads to diffuse hyperplasia in postnatal white matter. *J. Neurosci.* **28**, 914–922 (2008).
- Aguirre, A., Rubio, M. E. & Gallo, V. Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. *Nature* **467**, 323–327 (2010).
- Yuan, X. et al. Expression of the green fluorescent protein in the oligodendrocyte lineage: a transgenic mouse for developmental and physiological studies. *J. Neurosci. Res.* **70**, 529–545 (2002).
- Aguirre, A., Rizvi, T. A., Ratner, N. & Gallo, V. Overexpression of the epidermal growth factor receptor confers migratory properties to nonmigratory postnatal neural progenitors. *J. Neurosci.* **25**, 11092–11106 (2005).
- Hibbits, N., Pannu, R., John Wu, T. & Armstrong, R. C. Cuprizone demyelination of the corpus callosum in mice correlates with altered social interaction and impaired bilateral sensorimotor coordination. *ASN Neuro* **1**, e00013 (2009).
- Liebetanz, D. & Merkler, D. Effects of commissural de- and remyelination on motor skill behaviour in the cuprizone mouse model of multiple sclerosis. *Exp. Neurol.* **202**, 217–224 (2006).
- Schalomon, P. M. & Wahlsten, D. Wheel running behavior is impaired by both surgical section and genetic absence of the mouse corpus callosum. *Brain Res. Bull.* **57**, 27–33 (2002).
- Brooks, S. P. & Dunnett, S. B. Tests to assess motor phenotype in mice: a user's guide. *Nature Rev. Neurosci.* **10**, 519–529 (2009).
- Carter, R. J., Morton, J. & Dunnett, S. B. in *Current Protocols in Neuroscience* (John Wiley & Sons, 2001).
- Heimberger, A. B. et al. Brain tumors in mice are susceptible to blockade of epidermal growth factor receptor (EGFR) with the oral, specific, EGFR-tyrosine kinase inhibitor ZD1839 (Iressa). *Clin. Cancer Res.* **8**, 3496–3502 (2002).
- Jin, K. et al. Cerebral neurogenesis is induced by intranasal administration of growth factors. *Ann. Neurol.* **53**, 405–409 (2003).
- Illum, L. Transport of drugs from the nasal cavity to the central nervous system. *Eur. J. Pharm. Sci.* **11**, 1–18 (2000).
- Dhuria, S. V., Hanson, L. R. & Frey, W. H. Intranasal delivery to the central nervous system: mechanisms and experimental considerations. *J. Pharm. Sci.* **99**, 1654–1673 (2010).
- Cantarella, C., Cayre, M., Magalon, K. & Pascale, D. Intranasal HB-EGF administration favors adult SVZ cell mobilization to demyelinated lesions in mouse corpus callosum. *Dev. Neurobiol.* **68**, 223–236 (2008).
- Sugiarto, S. et al. Asymmetry-defective oligodendrocyte progenitors are glioma precursors. *Cancer Cell* **20**, 328–340 (2011).
- Kang, S. H., Fukaya, M., Yang, J. K., Rothstein, J. D. & Bergles, D. E. NG2⁺ CNS glial progenitors remain committed to the oligodendrocyte lineage in postnatal life and following neurodegeneration. *Neuron* **68**, 668–681 (2010).
- Zhang, Y. et al. Notch1 signaling plays a role in regulating precursor differentiation during CNS remyelination. *Proc. Natl Acad. Sci. USA* **106**, 19162–19167 (2009).
- Raymond, M., Li, P., Mangin, J.-M., Huntsman, M. & Gallo, V. Chronic perinatal hypoxia reduces glutamate-aspartate transporter function in astrocytes through the Janus kinase/signal transducer and activator of transcription pathway. *J. Neurosci.* **31**, 17864–17871 (2011).
- Bi, B. et al. Cortical glial fibrillary acidic protein-positive cells generate neurons after perinatal hypoxic injury. *J. Neurosci.* **31**, 9205–9221 (2011).

Acknowledgements We thank J. Dupree for advice on electron microscopy analysis. We thank L.-J. Chew, J. Corbin and J. Liu for comments on the manuscript. We thank M. McKenna for discussion on metabolic studies. We thank D. Bergles for providing the PDGFR-CreER mice and D. W. Threadgill for the EGFR^{fl/fl} mice. We thank R. Packer for his support. This work was supported by National Institutes of Health grants: K08NS073793 (J.S.); NSADA K12NS052159 (J.S.); K08NS069815 (S.S.); P01 NS062686 (V.G., T.L.H.); R01NS045702 (V.G.); R01NS056427 (V.G.); P30HD040677 (V.G.); R01MH067528 (F.H.); P30 NS05219 (F.H.); R01MH067528 (F.H.); and the Pioneer Award DP1 OD006850 (T.L.H.). The Childhood Brain Tumor Foundation (J.S.) and the National Brain Tumor Society (J.S.) also provided support.

Author Contributions J.S. designed all experiments with V.G. J.S. performed all experiments except electron microscopy and DTI imaging. T.R.H. performed all the Notch analyses. S.S. performed metabolic studies. J.R. and J.S. performed all electrophysiological analyses of CAPs. B.J. and M.R. assisted with experiments. K.S.-B. and T.L.H. performed the electron microscopy studies. D.C., Y.H. and F.H. performed DTI imaging and analysis. R.J.M. performed statistical analyses with J.S. on all behavioural experiments. V.G. supervised the entire project. J.S. and V.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.G. (vgallo@childrensnational.org).

METHODS

Animals. The CNP-EGFP (Rep) and CNP-EGFP-hEGFR (Rep-hEGFR) strains were generated as described previously and backcrossed to a C57BL/6 genetic background greater than nine generations^{4,11,12}. The Rep-hEGFR mice used for these experiments were crossed with heterozygote Rep mice to ensure that all pups were positive for GFP, but not all expressed hEGFR. This ensured littermate controls. For experiments that required only the Rep line of mice, a Rep adult mouse was crossed with a C57BL/6 adult mouse (Jackson Laboratories). Only mice that expressed eGFP (Rep) during screening on P2 with ultraviolet goggles were used. The PDGF α R-CreERT2 line (courtesy of D. Bergles; referred to in the text as PDGF α R-CreER) was crossed with Z/EG reporter mice (Jackson Laboratories; stock number 003920) and genotyped as previously reported²⁵. Z/EG reporter mice were crossed with EGFR^{fl/fl} (ref. 29) followed by PDGF α R-CreER transgenic mice. These mice were backcrossed to ensure homozygote EGFR^{fl/fl} mice, so that all mice were PDGF α R-CreER;EGFR^{fl/fl};Z/EG. To induce Cre recombination, tamoxifen was administered at a dosage of 75 μ g g⁻¹ body weight. The Ascl1^{GFP} mice (Jackson Laboratories, stock 012881; also known as Mash1) were used to determine the effects of gefitinib on white matter Ascl1-expressing cells. Unless described later or in the figure legends, only male mice were used, owing to male preterm children showing more clinically relevant injury and neurological deficits compared with females^{30,31}. All animal procedures were performed according to the Institutional Animal Care and Use Committee of the Children's National Medical Center and the *Guide for the Care and Use of Laboratory Animals* (National Institutes of Health).

Hypoxic rearing. Mice were randomly chosen to either undergo hypoxic rearing or serve as normoxia controls. The hypoxic mice were placed in a sealed chamber maintaining O₂ concentration at 10.5% by displacement with N₂ as described previously^{27,28,32}. Hypoxia began at P3 for 8 days until P11. This time window in rodent white matter oligodendrocyte development reproduces changes that occur at 23–40 weeks of gestation in the human brain⁹. A separate group of mice from the same breeding cage were used for age- and strain-matched normoxia controls. Genotyping (CNP-hEGFR; PDGF α R-CreER;Z/EG; Ascl1^{GFP}; EGFR^{fl/fl}; and Notch reporter mice) was performed at P11 by PCR of tissue obtained from the tail as previously reported^{4,11,25,29,33}. Time points chosen for immunohistochemistry or protein quantification were P11, P14.5, P15, P18, P30 and P60.

BrdU administration. The BrdU labelling protocol was performed in all mice as follows. Mice were injected intraperitoneally (i.p.) at the same time of the day with BrdU (50 μ g g⁻¹ body weight) daily for 4 days (P11–P14) in the morning. In studies using HB-EGF or gefitinib, BrdU was administered 1 h before vehicle or drug administration.

Gefitinib administration. Gefitinib (Iressa; Astra Zeneca (Tocris)) was prepared with strong sonication in 25% dimethylsulphoxide (DMSO) and 75% sunflower seed oil at a concentration of 10 mg ml⁻¹. Male mice in each litter were randomly chosen to receive either drug or vehicle. The drug dose for this study was 75 mg kg⁻¹ day⁻¹ and was administered once daily. An equal amount of vehicle was administered to control animals. A total of seven doses of vehicle or drug were administered beginning at P12.

HB-EGF administration. The intranasal route allows for small molecules to rapidly enter the cerebrospinal fluid from the nasal cavity, followed by subsequent distribution to the brain and spinal cord^{34–37}. Recombinant human heparin-binding epidermal growth factor constituent free (HB-EGF; R&D Biosciences) was prepared using 0.45% normal sterile saline solution at a concentration of 20 μ g ml⁻¹ and stored at –20 °C. Mice were randomized to the vehicle (saline) or HB-EGF group. Saline or HB-EGF was administered intranasally at no more than 5 μ l increments 5–10 min apart for a total of 100 ng g⁻¹. The mouse was held ventral-side up, and a small-modified 27-French catheter was inserted into either nare. Saline or drug was slowly administered and the mouse was held for 1–2 min to ensure absorption. Drug or saline was administered every 12 h beginning on the evening of P11.

Immunohistochemistry and antibodies. Freshly cut, free-floating brain sections (40 μ m thick) from P11–P60 mice were prepared as described previously^{3,11,27,32}. Primary antibody dilutions were 1:500 for anti-BrdU (Accurate), anti-NG2 (Millipore), anti-Olig2 (Millipore), anti-Ki67 (Vector), anti-APC (also referred as CC1; Millipore) and anti-cleaved caspase 3 (caspase 3; Millipore); 1:250 for anti-MBP (Covance); and 1:500 anti-EGFR phosphorylated Tyr 1068 (Novus Biologicals). Sections were incubated at room temperature (22–25 °C) for 1–3 h, followed by overnight at 4 °C in primary antibodies diluted in 0.1 M PBS containing 0.1% Triton X-100 and 5% normal goat serum (vol/vol). Three washes were performed with cold 1 PBS before secondary antibodies being administered. The secondary antibodies (1:200) used were AlexaFluor 488, AlexaFluor 546 and AlexaFluor 633 conjugated goat anti-rabbit, anti-rat or anti-mouse IgG (Invitrogen). Sections were incubated with secondary antibodies for 1 h at room temperature, followed by three 1 \times PBS washes. Sections were treated with 4',6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich) for 10–15 min at room temperature and mounted with Mowiol. Human tissue was

obtained from a deceased 3-day-old infant born at 36-week gestation with approval from the Children's National Institutional Review Board. Sections of corpus callosum and periventricular region were removed after fixation of the brain for 2 weeks in formalin solution. Tissue was kept in PBS for 1 week with daily changes of PBS solution to remove excess formalin. The brain was then placed in 20% glycerol solution for 24 h. Freshly cut, free-floating sections (40 μ m thick) were made on a sliding microtome. Sections were immunostained as described earlier.

Microscopy and cell counting. All fluorescent images were taken on a Zeiss LSM confocal laser-scanning microscope with sequential scanning mode using $\times 40$ oil objectives. Microscopy and cell counting were performed as recently described^{138,39}. Z-stack images of 1- μ m-thick single planes were captured throughout the entire thickness of the slice and each cell was analysed using a Zeiss LSM Image Browser (version 4.2) in its entire z-axis to exclude false double labelling due to overlay of signals from different cells. Four different laser lines were used to perform image localization of fluorescein isothiocyanate (FITC) (488 nm excitation; 522/35 emission filter), CY3 (560 nm excitation; 605/32 emission filter), Cy5 (647 nm excitation; 680/32 emission filter) and DAPI (400 nm excitation). Data acquisition and processing were controlled by modified LSM software. Analysis of immunofluorescence was performed on a confocal z-stack as previously described^{38,39}. Cells were counted in 225 \times 225 \times 10 μ m (X, Y, Z planes) images for cells per volume quantifications. Data were obtained from an average of six tissue sections per mouse per immunostain. Analysis of subcortical white matter—rostral to the hippocampus—was performed in the corpus callosum, cingulum and external capsule. Cells were counted in a blinded fashion. The merged image for each confocal z-stack was analysed and positive immunofluorescence identified for each individual channel. Merged images were processed in Photoshop CS5.5 with minimal manipulation of contrast.

Propidium iodide. To assess loss of membrane integrity and increased plasmalemma permeability *in vivo*, propidium iodide was used as demonstrated previously⁴⁰. Propidium iodide (10 mg ml⁻¹; Sigma) was diluted in 0.9% NaCl and 1 mg kg⁻¹ was administered i.p. 1 h before mice were killed. As described earlier, mice were perfused, brains were collected from all groups and free-floating brain sections (40 μ m thick) were prepared. Sections were washed, incubated with DAPI for 10 min, washed and then mounted on a slide. Propidium iodide (PI) emits bright red fluorescence when bound to RNA or DNA. Confocal microscopy was used to visualize the GFP⁺ PI⁺ DAPI⁺ cells.

Western blot analysis. For western blot analysis of white matter lysates, the subcortical white matter was dissected on ice-cold medium from 300–400- μ m-thick sections as previously described^{4,27,32}. Briefly, brains were sliced coronally and only sections anterior to the hippocampus were used. Using Roboz—a fine-straight and fine-angled microdissecting forceps under a dissecting microscope—the cortex was dissected away leaving the subcortical white matter attached to the striatum. The white matter was then easily pushed away from the striatum, leaving only white matter tissue. The dissected white matter was rinsed with ice-cold PBS, and then lysed on ice in 150–200 μ l of RIPA lysis buffer with protease inhibitors. Protein concentrations were determined by using the Bradford protein assay kit (Bio-Rad). Western blot analysis was performed using 10–40 μ g of total cell lysates. Proteins were resolved on 4–20% Tris glycine gels (NuSep) and transferred to Immobilon PVDF membranes in transfer buffer overnight at 4 °C. Membranes were blocked for 1 h in 5% milk in Tris-buffered saline-Tween 20 (TBST), then incubated at 4 °C overnight with primary antibodies diluted in 5% milk in TBST: 1:5,000 for anti-MBP (Covance), anti-CNP (Covance) and anti-actin (Millipore); 1:1,000 for anti-PLP (Abcam); 1:1,000 for anti-HB-EGF (Santa Cruz); 1:1,000 for anti-Delta1 (Santa Cruz); 1:1,000 for anti-NICD (Iowa Hybridoma Bank C17.9C6); and 1:4,000 for anti-aspartoacylase (ThermoScientific). The membranes were then washed in TBST three times for 10–15 min at room temperature followed by the addition of either horseradish-peroxidase-conjugated goat polyclonal anti-rabbit IgG (Santa Cruz) for polyclonal primary antibodies, or horseradish-peroxidase-conjugated goat anti-mouse (Santa Cruz) for mouse monoclonal primary antibodies diluted in 5% milk in TBST. For phosphorylated EGFR (Novus Biologicals), 5% bovine serum albumin (BSA) in TBST was used as a block and for primary antibody incubation overnight. For all western blots, chemiluminescent signals were detected using Pierce ECL western blotting substrate. X-ray films were scanned using an Agfa T1200 scanner and densitometric measurements were obtained using ImageJ software (<http://rsb.info.nih.gov/ij/>). Western blots were obtained from the white matter of 3–6 male mice in each group. Densitometric measurements were obtained using ImageJ software averaged as previously described^{4,27,32}.

EGF ELISA. White matter was grossly dissected on ice-cold 1 \times PBS from P11, P15 and P18 Rep and Rep-hEGFR normoxia and hypoxia mice as described earlier. Assay procedure was performed according to the manufacturer's instructions (R&D Systems, Mouse EGF Quantikine ELISA Kit). Experiments were performed in triplicates and averaged.

FACS. White matter microdissected tissue from hypoxia-exposed P15 Rep⁺ (CNP-EGFP⁺) vehicle- or HB-EGF-treated mice were FACS purified. Two-to-three male and female brains were pooled for each sample (individual *n*). Tissue was dissociated into single-cell suspensions as previously described^{11,41}, and analysed for light forward and side scatter using a FACS Aria instrument (BD Bioscience). The collected Rep⁺ cells were used for western blot analysis.

Electron microscopy. Mice at P30 and P60 were perfused with 4% paraformaldehyde containing 10% picric acid and 5% glutaraldehyde and post-fixed for 2 weeks⁴². Brains were sectioned and prepared in groups at the same time as previously described^{32,42}. Sagittal sections of white matter were examined with a JEOL transmission electron microscope (JEM-1400), and pictures were taken with a Gatan SC1000 ORIUS CCD camera. Measurements and image processing was performed using ImageJ. Myelin thickness was calculated from the average of radial measurements at four points per sheath, avoiding areas of tongue processes or fixation artefact^{42,43}. Axon diameters were calculated from measurement of the axon circumference. Axons with diameters typical of unmyelinated fibres (<0.3 µm) were excluded from analysis^{42,43}. The extent of myelination was quantitatively compared by determining *g* ratios, which were calculated by dividing the diameter of the axon by the diameter of the entire myelinated fibre, as previously described^{42–44}. Measurements were performed by J.S., who was blinded to groups. At least 100 axons were measured for each brain.

LPC-induced demyelination. Bilateral demyelination was performed in adult male and female C57BL/6 (8 weeks old) mice after deep ketamine/xylazine anaesthesia (10 mg g⁻¹ body weight). Mice were placed in a modified stereotaxic frame (Stoetling) and 2 µl of 2% lysolecithin (EMD Chemicals, LPC) solution (vol/vol) and/or 0.9% NaCl (vol/vol) was injected bilaterally into the corpus callosum using a Hamilton micropipette (Stoetling). Injection time lasted for 5 min to reduce reflux along the needle track. The needle was then slowly withdrawn over a 5 min period. Stereotaxic coordinates for the corpus callosum were taken from Bregma (0.26 mm caudal, 1.0 mm lateral and 3.0 mm ventral). Inclined-beam walking behavioural test testing began 5 days after surgery in both groups of mice as described later. Bilateral demyelination was confirmed after testing by perfusing the mice as described earlier and immunohistochemical analysis of the corpus callosum was performed using anti-MBP. Only mice that displayed clear bilateral lesions on microscopic examination were included in the behavioural analysis. Three (*n* = 3) mice were excluded from the study because bilateral white matter demyelination was not clearly evident.

DTI analysis. Mice (P60) used for DTI were perfused and imaged as previously described⁴⁵. One hour before DTI scans the brains were soaked 3 times for 10 min each time in 10 ml PBS to remove the PFA solution. The brains were placed into a custom-built MRI-compatible tube filled with Fluorinert—an MRI susceptibility-matching fluid (Sigma-Aldrich). The DTI data sets were obtained on a 9.4 T horizontal bore magnet (Bruker) with a custom-made ¹H radio frequency coil. The DTI experiments were performed using the Stejskal–Tanner spin-echo diffusion-weighted sequence with a diffusion gradient of 5 ms and a delay between the two diffusion gradients of 15 ms. Twenty-four contiguous coronal slices of 0.5 mm thickness were acquired using a repetition time (TR) of 2 s and an echo time (TE) of 25.1 ms. Two Shinnar–Le Roux (SLR) pulses of 1 ms each were used for excitation and inversion, respectively. Twenty averages were acquired for each slice and the 128 × 64 pixel resolution images were zero-filled to 256 × 256 pixel resolution, resulting in an in-plane spatial resolution of 100 µm × 100 mm². Sixteen different images were acquired for each slice, 15 corresponding to various non-collinear diffusion-weighting directions with *b* = 1,000 s mm⁻² and one with no diffusion weighting. DTI processing and analysis was performed blindly as described previously^{45,46}.

CAPs. Compound action potential (CAP) recordings were performed in all four groups at P30 and P60 using methods previously described^{4,39}. Briefly, after mice were killed, coronal slices 400 µm in thickness were obtained using a VT1000S vibratome (Leica) in ice-cold slicing solution (85 mM NaCl, 2.5 mM KCl, 25 mM NaHCO₃, 1.25 mM NaH₂PO₄, 0.5 mM CaCl₂, 7 mM MgCl₂, 25 mM glucose, 75 mM sucrose). Slices were placed in recording solution (125 mM NaCl, 2.5 mM KCl, 26 mM NaHCO₃, 1.25 mM NaH₂PO₄, 2.5 mM CaCl₂, 1.3 mM MgCl₂, 11 mM glucose, pH 7.4) bubbled in 95% O₂ and 5% CO₂, and maintained at 37 °C for 1 h then kept in the same solution at room temperature until recording. CAP recordings were performed on five slices corresponding to sections 13–18 (Bregma 1.32 to –0.94 mm) of the National Institutes of Mental Health's mouse brain atlas (http://www.mbl.org/atlas170/atlas170_frame.html). Slices were placed in a recording chamber superfused with oxygenated recording solution at a flow rate of 2 ml min⁻¹ and viewed using the ×10 objective of an Olympus BX61WI microscope. CAP recordings were obtained using an FHC concentric bipolar stimulating electrode and an extracellular field electrode with a tip resistance of 1 MΩ. The stimulating and recording electrodes were placed ~2 mm apart (~1 mm on each side of the midline) in the corpus callosum of subcortical white matter, and a constant stimulus was delivered for each recording in current clamp. A single pulse protocol was

used with a pulse frequency of 200 Hz and a period of 5 ms. Fifty sweeps were recorded, averaged and used in the analysis for CAP amplitude. Two distinguishable downward waves were evident for each CAP recording, with the first corresponding to rapidly propagating myelinated (M) axons and the second to slower propagating unmyelinated (UM) axons^{4,39,47}.

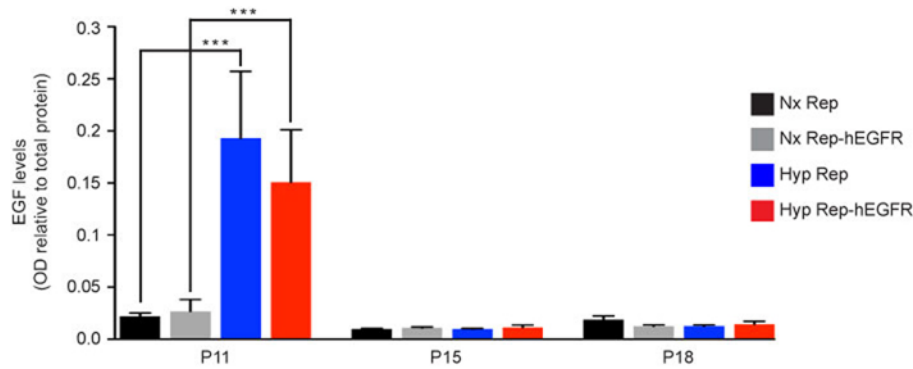
¹H-NMR. Brains from all groups were collected at P11, P18 and P30. The brains were removed and placed on dry ice. The white matter was dissected, snap frozen in liquid nitrogen (total time <60 s), and stored in a –80 °C freezer until extraction. Samples for ¹H-NMR spectroscopy were prepared as previously published⁴⁸. Each frozen sample was homogenized in 1 ml of 7% perchloric acid and centrifuged for 10 min at 4 °C and 4,600g. Supernatants were transferred to separate tubes and pellets were re-extracted. Combined supernatants were neutralized with KOH, centrifuged and lyophilized. Lyophilized samples were dissolved in 0.8 ml of 99% D₂O and pH was adjusted to 7.0. Fully relaxed ¹H-NMR spectra were acquired on Varian 500 with the following parameters: 90° pulse angle, an acquisition time of 1.36 s, relaxation delay of 17 s, total number of 800 scans per sample. Low-power pre-saturation pulse at water frequency was used to achieve water suppression. Obtained spectra were analysed using MestReNova software (version 8.1; Mestrelab Research) and the amounts of metabolites were quantified from integrals of the peak areas corrected for number of protons and using 2,2,3,3-D(4)-sodium-3-trimethylsilylpropionate as an internal control.

Behavioural testing. Each behavioural experiment was performed in separate naive mice that had not undergone any previous behavioural testing. The complex running wheel task was performed as previously described^{14–16}. At P45, naive mice that had not undergone any previous behavioural testing were individually housed in a modified cage equipped with a running wheel attached to an optical sensor to constantly detect the number of wheel revolutions per time interval (minute). Animals were kept on a regular 12 h light/dark cycle. Food and water were made available *ad libitum*. During the first 2 weeks, a training wheel with all 38 rungs was present, allowing for normalization of running behaviour. On the third week (day 15; age P60), the regular training wheel was replaced with a complex wheel of the same diameter with 22 rungs missing in an alternative pattern. Using Activity Wheel Monitoring Software (Lafayette Instruments), wheel revolutions were recorded each day and exported to a Microsoft Excel file in which daily total distance travelled and maximum daily velocity were calculated. All mice showed spontaneous running behaviour and no mice were excluded from this study. The inclined beam-walking task was performed as previously described^{17,18}. Two elevated 80 cm in length wooden beams were placed at a 30° angle. One beam was 2 cm in width and the other was 1 cm in width. A dark box with bedding was at the end of the incline and served as a target for the mouse to reach. A blinded experimenter observing and recording from above assessed mouse performance by documenting the number of foot slips (either hind legs or front legs) and the time to traverse the beam¹⁸. In pilot studies, we determined that hypoxia-exposed Rep mice less than 30 days of age—or on a beam that was inclined more than 30°—were unable to perform this task (data not shown). To confirm whether this sensorimotor task is dependent on subcortical white matter, bilateral LPC- or 0.9% saline-injected adult mice were tested on day 5 after surgery (described earlier). Bilateral demyelination was confirmed after testing by removal of brains and immunohistochemical analysis of corpus callosum. Only mice that had clear bilateral lesions on microscopic examination were included in the behavioural analysis (*n* = 3 mice were excluded due to failure in demyelination).

Statistics. All data in histograms are presented as averages ± s.e.m. All cell counting and western blot data were statistically compared using one-way ANOVA to determine whether overall differences exist across study groups. Comparisons between specific groups were treated as unplanned comparisons, which were adjusted using a Bonferroni correction. A two-tailed type 1 error (*P* value < 0.05) was used to determine statistical significance. Each experiment evaluated outcomes in four groups. The Bonferroni correction was applied for the following comparisons: normoxia (Nx)_a versus Nx_b; Nx_a versus hypoxia (Hyp)_a; and Nx_b versus Hyp_b. The two hypoxia groups were compared post-hoc if the one-way ANOVA was significant using a two-tailed unpaired *t*-test (Hyp_a versus Hyp_b) with two-tailed type 1 error set at *P* = 0.05. For the electron microscopy data, a one-way ANOVA was used to compare the *g* ratio for each mouse in each of the four respective groups. If significance was found, then a two-tailed unpaired *t*-test with two-tailed type 1 error set at *P* = 0.05 was used to make the following comparisons: Nx_a versus Hyp_a; Hyp_a versus Hyp_b; Nx_a versus Nx_b; and Nx_b versus Hyp_b. Scatter plots of *g* ratios of individual fibres in relation to axon diameter are shown comparing the groups outlined earlier. The average FA anisotropy for each mouse in each group was compared using a one-way ANOVA as described earlier. The Bonferroni correction was applied for the following comparisons: Nx saline versus Hyp saline; Nx saline versus Hyp HB-EGF; and Hyp saline versus Hyp HB-EGF. For the complex running wheel data, we used a longitudinal linear regression analysis to compare slopes (trajectories) of change. We also performed post-hoc testing

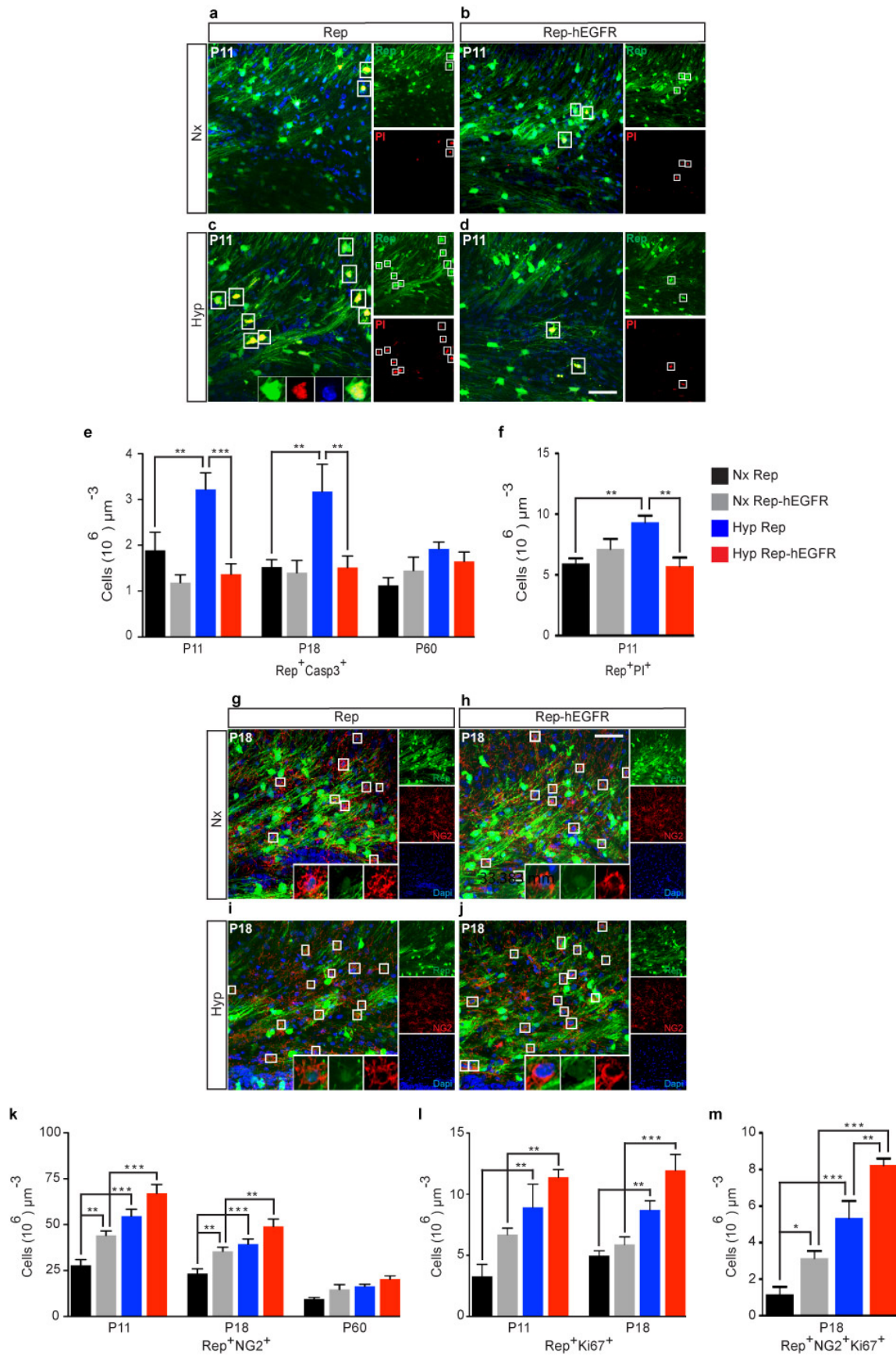
between the four groups for each individual day. For the beam-walking behavioural results, the number of foot slips, the time to traverse the beam and the size of the beam were analysed using a Poisson multiple regression analysis, which allowed us to overcome the lack of normality in count-type data and account for other variables to compare the rate of foot slips. The one-way ANOVA with post hoc comparisons was performed using GraphPad Prism 5.0 (for Mac). All histograms and scatter plots in this manuscript were created with GraphPad Prism.

29. Lee, T.-C. & Threadgill, D. W. Generation and validation of mice carrying a conditional allele of the epidermal growth factor receptor. *Genesis* **47**, 85–92 (2009).
30. Ment, L. R. *et al.* Change in cognitive function over time in very low-birth-weight infants. *J. Am. Med. Assoc.* **289**, 705–711 (2003).
31. Ment, L. R. *et al.* Longitudinal brain volume changes in preterm and term control subjects during late childhood and adolescence. *Pediatrics* **123**, 503–511 (2009).
32. Jablonska, B. *et al.* Oligodendrocyte regeneration after neonatal hypoxia requires FoxO1-mediated p27Kip1 expression. *J. Neurosci.* **32**, 14775–14793 (2012).
33. Mizutani, K. *et al.* Differential Notch signalling distinguishes neural stem cells from intermediate progenitors. *Nature* **449**, 351–355 (2007).
34. Born, J. *et al.* Sniffing neuropeptides: a transnasal approach to the human brain. *Nature Neurosci.* **5**, 514–516 (2002).
35. Thorne, R. G., Pronk, G. J., Padmanabhan, V. & Frey, W. H. Delivery of insulin-like growth factor-I to the rat brain and spinal cord along olfactory and trigeminal pathways following intranasal administration. *Neuroscience* **127**, 481–496 (2004).
36. Ross, T. M. *et al.* Intranasal administration of interferon beta bypasses the blood-brain barrier to target the central nervous system and cervical lymph nodes: a non-invasive treatment strategy for multiple sclerosis. *J. Neuroimmunol.* **151**, 66–77 (2004).
37. Fletcher, L. *et al.* Intranasal delivery of erythropoietin plus insulin-like growth factor-I for acute neuroprotection in stroke. *J. Neurosurg.* **111**, 164–170 (2009).
38. Ishibashi, N. *et al.* White matter protection in congenital heart surgery. *Circulation* **125**, 859–871 (2012).
39. Ritter, J. *et al.* Neonatal hyperoxia exposure disrupts axon-oligodendrocyte integrity in the subcortical white matter. *J. Neurosci.* **33**, 8990–9002 (2013).
40. Whalen, M. J. *et al.* Acute plasmalemma permeability and protracted clearance of injured cells after controlled cortical impact in mice. *J. Cereb. Blood Flow Metab.* **28**, 490–505 (2008).
41. Aguirre, A. & Gallo, V. Postnatal neurogenesis and gliogenesis in the olfactory bulb from NG2-expressing progenitors of the subventricular zone. *J. Neurosci.* **24**, 10530–10541 (2004).
42. Furusho, M., Dupree, J. L., Nave, K.-A. & Bansal, R. Fibroblast growth factor receptor signaling in oligodendrocytes regulates myelin sheath thickness. *J. Neurosci.* **32**, 6631–6641 (2012).
43. Zhou, Y.-X., Pannu, R., Le, T. Q. & Armstrong, R. C. Fibroblast growth factor 1 (FGFR1) modulation regulates repair capacity of oligodendrocyte progenitor cells following chronic demyelination. *Neurobiol. Dis.* **45**, 196–205 (2012).
44. Liu, J. *et al.* Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nature Neurosci.* **15**, 1621–1623 (2012).
45. Duque, A. *et al.* Neuroanatomical changes in a mouse model of early life neglect. *Brain Struct. Funct.* **217**, 459–472 (2012).
46. Chahboune, H. *et al.* Hypoxic injury during neonatal development in murine brain: correlation between *in vivo* DTI findings and behavioral assessment. *Cereb. Cortex* **19**, 2891–2901 (2009).
47. Crawford, D. K., Mangiardi, M. & Tiwari-Woodruff, S. K. Assaying the functional effects of demyelination and remyelination: revisiting field potential recordings. *J. Neurosci. Methods* **182**, 25–33 (2009).
48. Scafidi, S. *et al.* Metabolism of acetyl-L-carnitine for energy and neurotransmitter synthesis in the immature rat brain. *J. Neurochem.* **114**, 820–831 (2010).



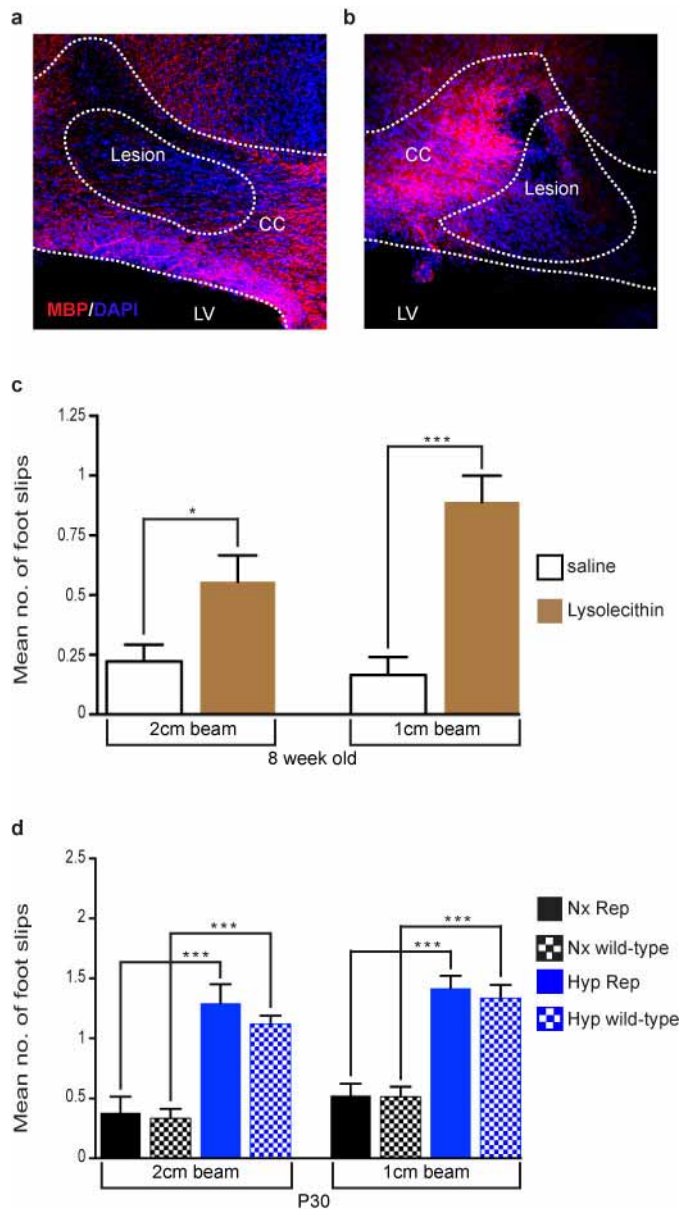
Extended Data Figure 1 | Hypoxia results in a significant increase in EGF levels in the white matter. The white matter was dissected out at P11, P15 and P18 in normoxia (Nx)- and hypoxia (Hyp)-exposed CNP-EGFP (Rep) and CNP-EGFP-hEGFR (Rep-hEGFR) mice. At P11, in both Hyp groups, there was a significant increase in EGF levels, as measured by ELISA. There was no significant difference between the two Hyp groups, indicating that

overexpression of EGFR in oligodendrocyte lineage cells does not modify endogenous EGF levels. At P15 and P18, there was no difference between all four groups. All histograms are presented as mean absorption (optical density (OD)) relative to total protein concentration \pm s.e.m. $**P < 0.05$; $***P < 0.01$ (P11 and P15, $n = 4$ mice per group and per age; P18, $n = 3$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons).



Extended Data Figure 2 | Enhanced EGFR expression in oligodendrocyte lineage cells prevents oligodendrocyte death and promotes proliferation of OPCs in white matter. **a–d**, Representative $\times 40$ confocal images of Rep⁺PI⁺ cells from normoxia and hypoxia-exposed white matter at P11 in Rep and Rep-hEGFR mice. **e**, At P11 and P18, hypoxia resulted in a significant increase in the number of oligodendrocyte cells undergoing apoptosis (Rep⁺Casp3⁺). Enhanced EGFR expression prevented this increase at all time points, except P60 where no difference was evident. Comparison of the Hyp Rep with the Hyp Rep-hEGFR groups demonstrates that hEGFR in oligodendrocyte lineage cells is protective against apoptosis induced by hypoxia ($n = 4$ mice per group and per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **f**, PI was injected intraperitoneally 1 h before mice were killed. A significant increase in the number of Rep⁺PI⁺ oligodendrocyte cells indicated membrane disruption contributing to cell death ($n = 3$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **g–j**, Representative $\times 40$ confocal images of Rep⁺NG2⁺ OPCs from normoxia and hypoxia-exposed white matter at P18 in Rep and

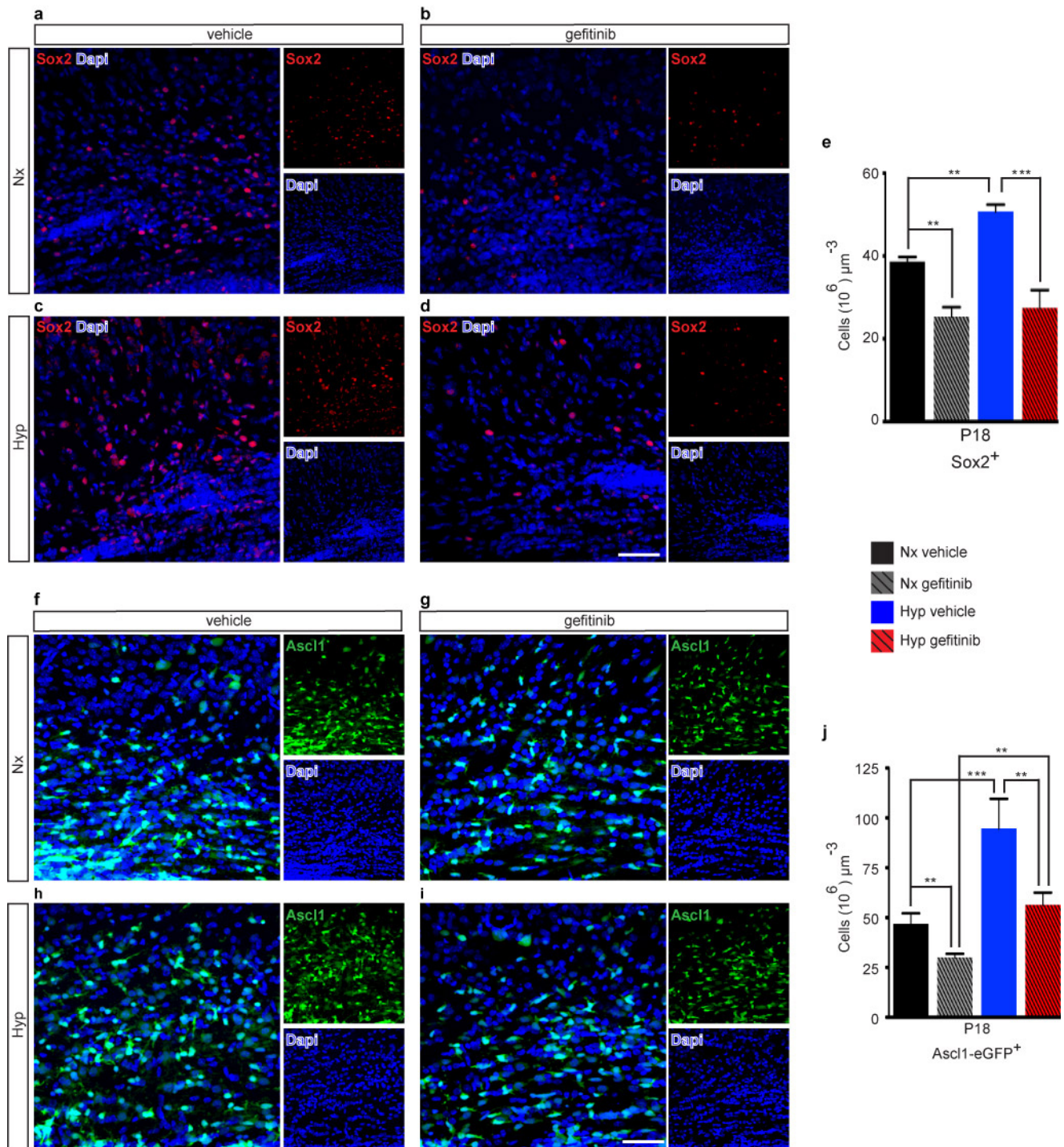
Rep-hEGFR mice. **k**, At P11 and P18, Rep-hEGFR Nx mice had more NG2⁺ OPCs compared with the Rep Nx group. Hypoxia resulted in a significant increase in the number of Rep⁺NG2⁺ OPCs in both Rep and Rep-hEGFR groups; however, overexpression of hEGFR did not have an additive effect. **l**, At P11 and P18, the Rep-hEGFR group had more Rep⁺Ki67⁺ cells; however, this did not reach significance ($P > 0.05$). Hypoxia resulted in enhanced oligodendrocyte-lineage proliferation at P11 and P18, but hEGFR overexpression did not have a significantly additive effect when compared with the Rep Hyp group. **k, l**, $n = 4$ mice per group and per age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons. **m**, More OPCs were in a proliferative state in Rep-hEGFR Nx mice compared with Rep Nx. Hypoxia enhanced OPC proliferation, and overexpression of hEGFR resulted in a significantly additive increase compared to the Rep Hyp group ($n = 3$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). All histograms are presented as means \pm s.e.m. Scale bars, 50 μ m (**a–d, g–j**). * $P = 0.05$; ** $P < 0.05$; *** $P < 0.01$.



Extended Data Figure 3 | The inclined beam-walking task requires normal subcortical white matter.

a–c, This study was performed to test the hypothesis that the inclined beam-walking task is a good assessment of subcortical white matter function. To test this hypothesis, we tested a well-established model of subcortical white matter demyelination induced by bilateral lysolecithin injections in 8-week-old C57BL/6J male and female mice. Animals were tested at 5 days after surgical intervention—which is a time period when demyelination is at its maximum—to determine whether subcortical white matter integrity is important in this behavioural task. Control mice received bilateral injections of normal saline using the same coordinates as the lysolecithin group. **a**, **b**, Bilateral demyelination was confirmed after testing by removal of brains and immunohistochemical analysis of corpus callosum. Only mice that had clear bilateral lesions on microscopic examination were included in the behavioural analysis ($n = 3$ mice were excluded). **c**, The lysolecithin-injected mice had a marginally significant or very significant increase in average number of foot slips on the 2-cm- and 1-cm-wide inclined beam, respectively (two-tailed Mann–Whitney test, $n = 6$ per group).

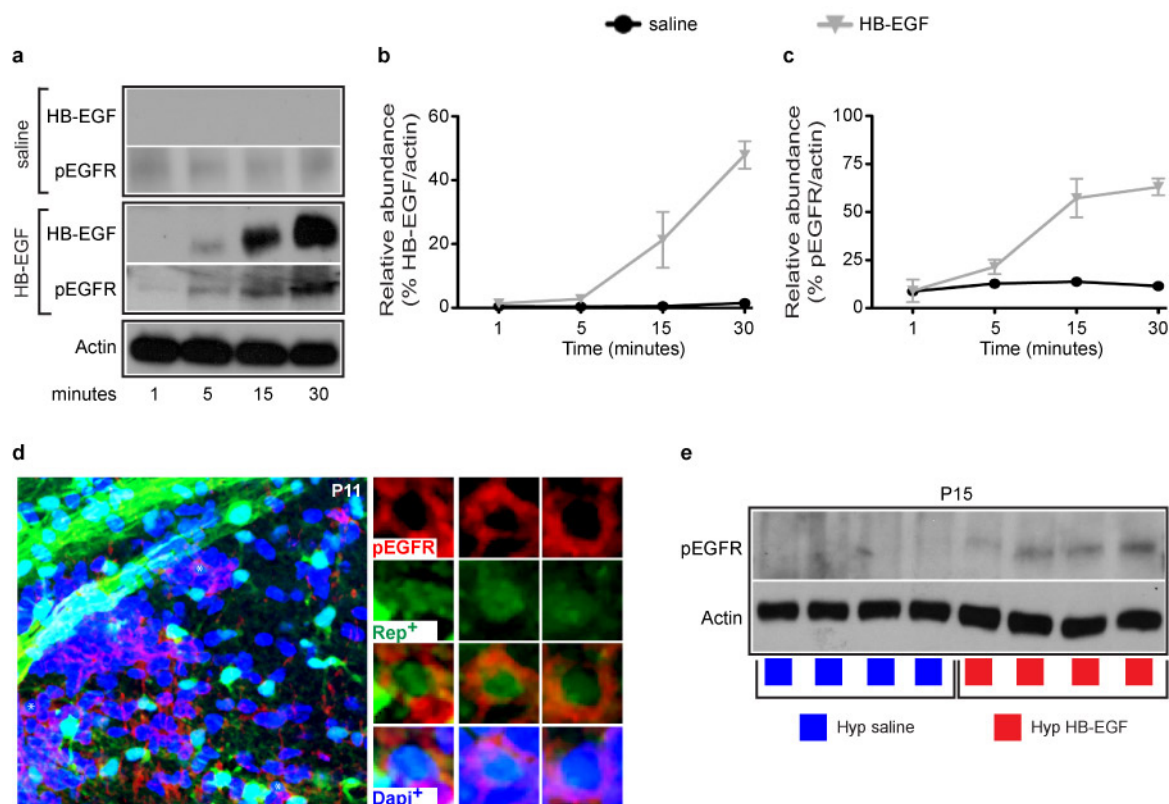
d, We wanted to determine whether the performance of Rep (CNP-eGFP) transgenic mice on the inclined beam-walking task was similar to that of C57BL/6 mice (wild type). No difference in performance was evident between the two different lines of male mice. We found that the Rep and wild-type mice performed similarly in either normoxic or hypoxic conditions (Poisson multiple regression analysis; Nx Rep, $n = 7$; Nx wild type, $n = 11$; Hyp Rep, $n = 8$; Hyp wild type, $n = 11$). All histograms are presented as means \pm s.e.m. * $P = 0.05$; ** $P < 0.05$; *** $P < 0.01$.



Extended Data Figure 4 | Inhibition of EGFR prevents expansion of progenitor cells in the developing white matter and after hypoxia.

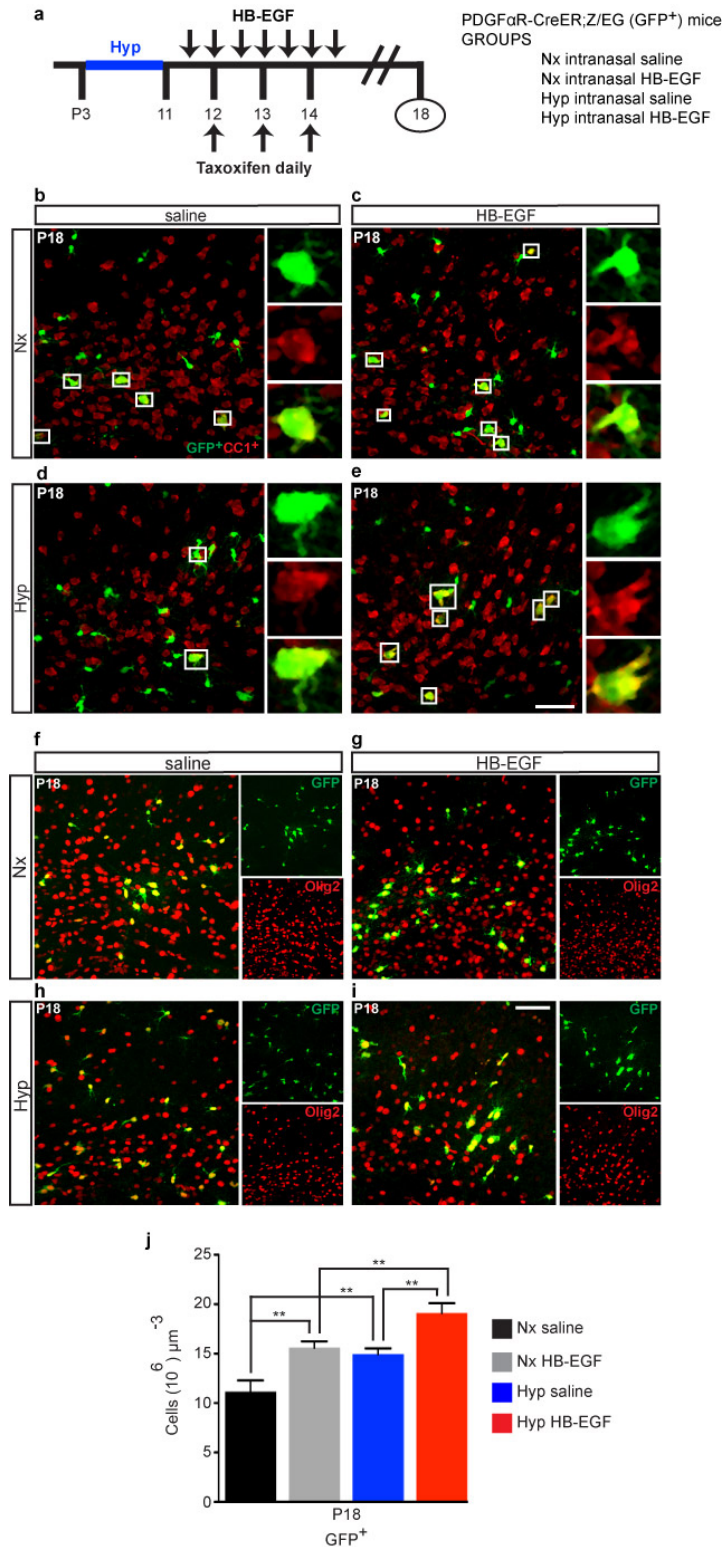
a–d, Representative $\times 40$ confocal images of subcortical white matter Sox2⁺ cells, a transcription factor expressed in proliferating multipotential neural progenitor cells. **e**, Gefitinib, a specific EGFR inhibitor, resulted in a significant decrease in the number of white matter Sox2-expressing cells compared with normoxia vehicle-treated mice. After hypoxia, there was a significant expansion of Sox2⁺ cells in the white matter compared with the Nx vehicle group; however, this expansion was prevented by gefitinib (Hyp gefitinib) (Hyp vehicle versus Hyp gefitinib, $P < 0.01$) ($n = 4$ for each group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons).

f–i, Representative $\times 40$ confocal images of subcortical white matter Ascl1⁺ cells (Mash1) in the Ascl1-eGFP transgenic mice. Ascl1 is a proneural transcription factor expressed in proliferating multipotential neural progenitor cells. **j**, Similar to above, gefitinib resulted in a significant decrease in the total number of Ascl1-eGFP⁺ cells in the white matter compared with normoxia vehicle-treated mice. Hypoxia resulted in a significant expansion in the number of Ascl1-eGFP⁺ cells, which gefitinib prevented (Hyp vehicle versus Hyp gefitinib, $P < 0.05$) ($n = 4$ for each group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). All histograms are presented as means \pm s.e.m. $**P < 0.05$; $***P < 0.01$. **a–d**, **f–i**, Scale bars, 50 μm .



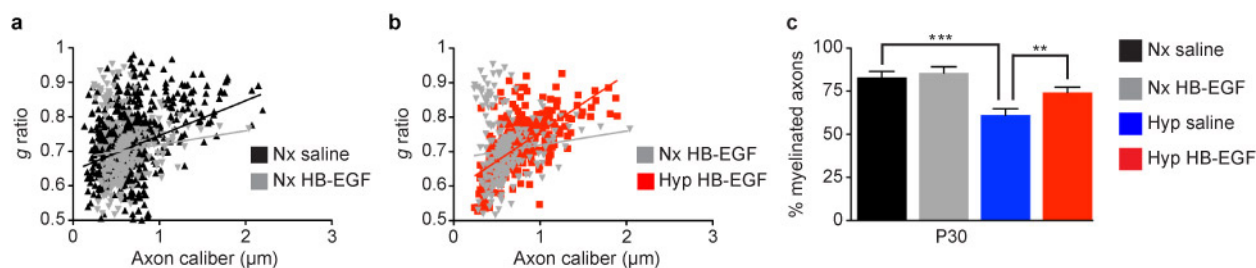
Extended Data Figure 5 | Intranasal HB-EGF does enter the brain and activates EGFRs in oligodendrocyte lineage cells. **a–c**, Saline or HB-EGF was administered intranasally once in P11 mice, which were then killed at 1, 5, 15 and 30 min after administration. **a**, Western blot analysis was performed on microdissected white matter probing for actin, HB-EGF and pEGFR (Tyr 1068 phosphorylation site). In the saline group, no HB-EGF was detected in the white matter and no change in pEGFR was detected. In the mice that received HB-EGF, the HB-EGF protein was detected at 5 min and increased up to 30 min. The pEGFR signal steadily increased at 5, 15 and 30 min after HB-EGF administration. **b, c**, The line graphs represent relative abundance of protein compared with actin ($n = 3$ for each time point and condition). Line graphs are presented as means \pm s.e.m. **d**, A normoxia-treated P11 mouse was

administered HB-EGF and killed 30 min later. Immunohistochemistry of pEGFR was performed. In the white matter, there were several Rep⁺pEGFR⁺DAPI⁺ cells, indicating that oligodendrocyte lineage cells express activated EGFR (pEGFR). Shown is a $\times 40$ representative image of the white matter demonstrating Rep⁺pEGFR⁺DAPI⁺ cells. **e**, In this set of experiments, Hyp Rep mice received either intranasal saline or HB-EGF from P11–P14 (Fig. 3b). The subcortical white matter was microdissected at P15 and CNP-eGFP⁺ cells were FACS purified. Western blot analysis was performed to probe for pEGFR. The western blot demonstrates that a more robust signal for pEGFR was present in the Hyp HB-EGF group, relative to actin ($n = 4$ for each group of 2–3 pooled brains).



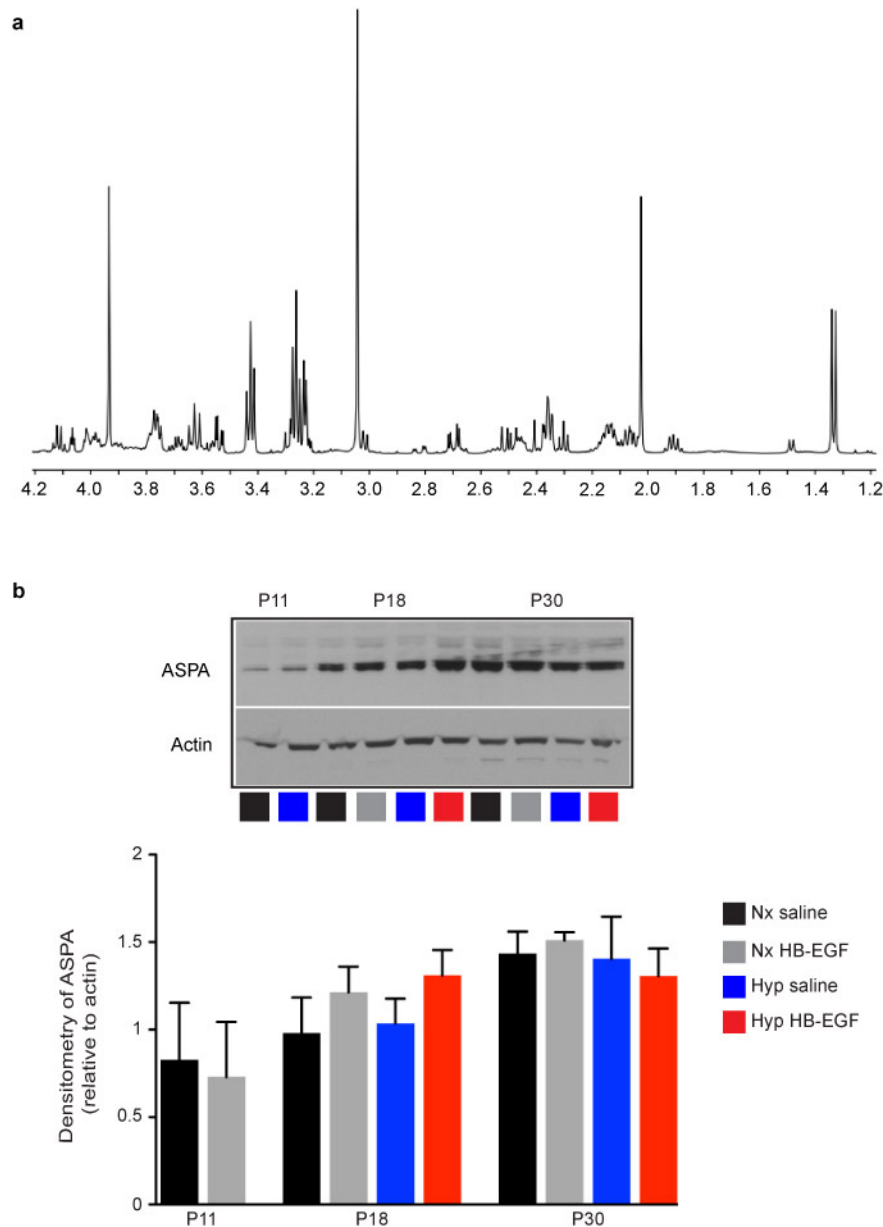
Extended Data Figure 6 | Intranasal HB-EGF treatment increases the number of oligodendrocyte lineage cells derived from PDGF α R-expressing OPCs. **a**, PDGF α R-CreER;Z/EG transgenic mice were divided into four groups. Saline or HB-EGF was administered intranasally. Intraperitoneal injections of tamoxifen were administered at P12, P13 and P14 in the morning, 1 h before the morning dose of saline/HB-EGF. Mice were killed at P18. **b–e**, Representative $\times 40$ confocal images of CC1 $^{+}$ cells (red) derived from PDGF α R-CreER;Z/EG (GFP $^{+}$) (green) progenitors at P18. **f–i**, Cells derived from PDGF α R-expressing progenitors after hypoxia and after HB-EGF treatment belong to the oligodendrocyte lineage (Olig2 $^{+}$). Representative $\times 40$

confocal images of the subcortical white matter in all four groups at P18. In serial sections from each PDGF α R-CreER;Z/EG mouse, all GFP $^{+}$ cells in the subcortical white matter co-stained with anti-Olig2 $^{+}$ antibody in all four groups ($n = 4$ for each group). In all four groups, no GFP $^{+}$ cells co-stained with anti-GFAP or anti-glutamate/aspartate transporter (GLAST) antibody (data not shown). **j**, Hypoxia results in a significant increase in the number of GFP $^{+}$ cells in the white matter and HB-EGF has an additive effect ($n = 4$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). Histograms are presented as means \pm s.e.m. **b–i**, Scale bars, 50 μ m. ** $P < 0.05$; *** $P < 0.01$.



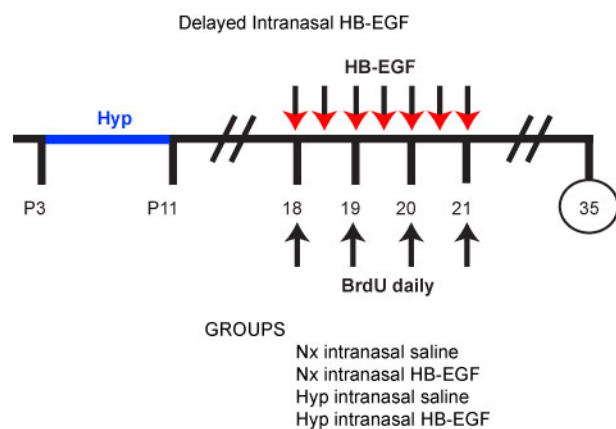
Extended Data Figure 7 | HB-EGF treatment prevents hypoxia-induced changes in white matter axonal *g* ratios at P30. **a, b,** Scatter plots depicting *g* ratios versus axon diameter. The lines represent linear fits to pooled data from all mice for each genotype ($n = 3$ mice per group). **a,** The scatter plot demonstrates that the Nx saline and Nx HB-EGF groups were similar.

b, The scatter plot demonstrates that the Nx HB-EGF and Hyp HB-EGF groups were similar. **c,** Histogram demonstrating that, at P30, the percentage of myelinated subcortical white matter fibres was significantly decreased in the Hyp saline group. No significant difference was found in the Hyp HB-EGF group. ** $P < 0.05$; *** $P < 0.01$.

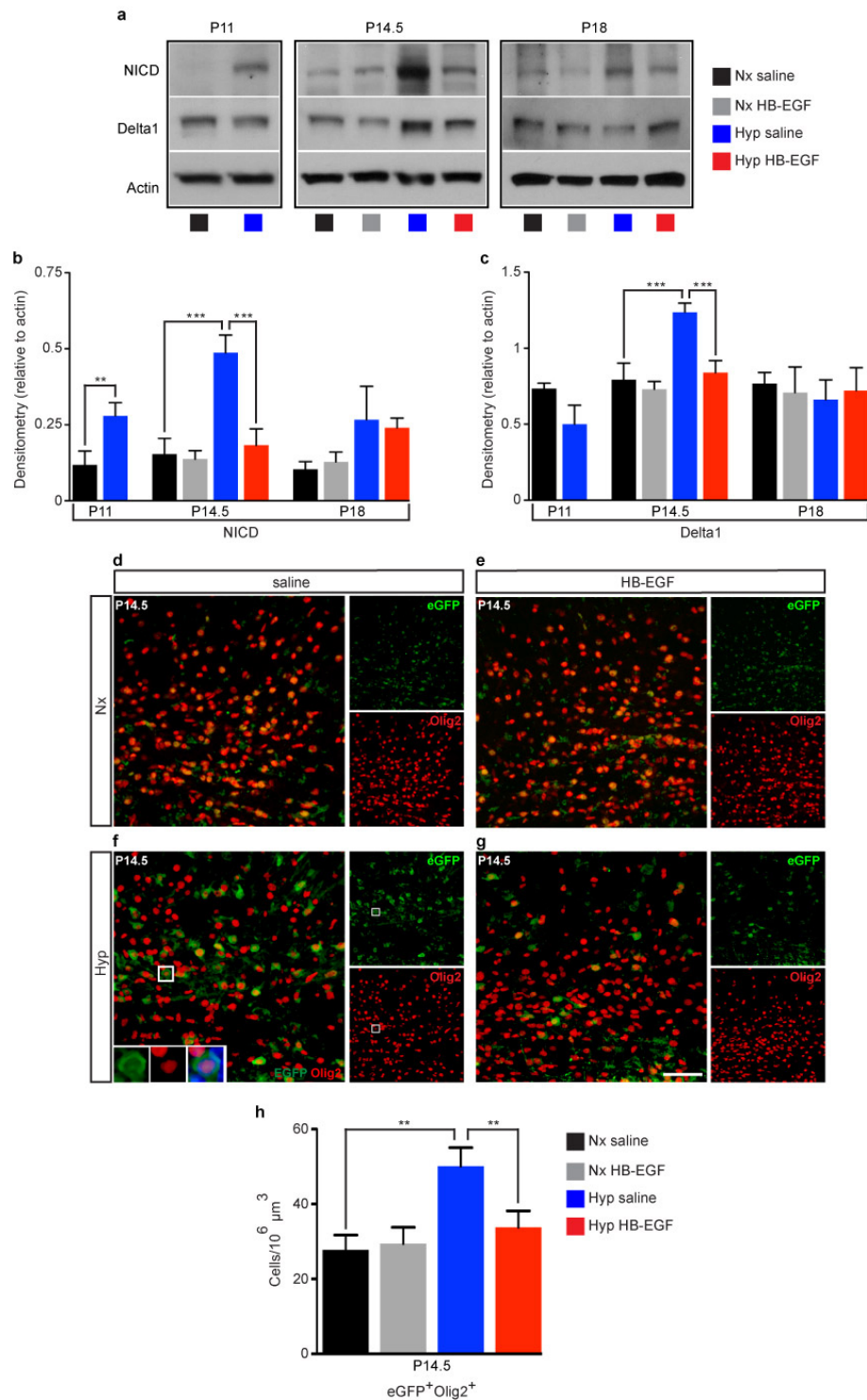


Extended Data Figure 8 | Intranasal HB-EGF prevents loss of NAA after hypoxia. **a**, ^1H -NMR spectroscopy was performed on dissected white matter at P11, P18 and P30. A full-scale representative spectrum is shown where the peak for NAA is at 2.0 p.p.m. The spectrum shown on Fig. 4l is truncated. **b**, Western blots of aspartoacylase (ASPA), an enzyme found in oligodendrocytes and responsible for hydrolysis of NAA for myelin production in the developing

brain. Hypoxia does not result in any significant change in the amount of ASPA present in the white matter at each of the time points listed (P11 and P30, $n = 4$ for each group and age; P18, $n = 5$ for each group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons, except P11 unpaired t -test). Histograms are presented as means \pm s.e.m.



Extended Data Figure 9 | Protocol for late HB-EGF administration used for this study. In this study, HB-EGF or saline was administered at a later time point. Beginning in the morning of P18, HB-EGF or saline was administered every 12 h until the morning of P21. Intraperitoneal BrdU was administered 1 h before HB-EGF or saline administration, from P18–P21. The inclined beam-walking task was performed at P35. Only Rep mice were used for this study. The histogram is presented in Fig. 4p.



Extended Data Figure 10 | Intranasal HB-EGF accelerates oligodendrocyte maturation in the white matter after hypoxia by preventing Notch activation. **a**, Microdissected white matter was probed for activated Notch intracellular domain (NICD) and its ligand Delta1. Western blot analysis obtained from microdissected white matter at P11, P14.5 and P18 with actin as a loading control. **b**, **c**, Histograms represent quantification of the density of NICD (**b**) and Delta1 (**c**) signal normalized to actin. **b**, At P11 and P14.5, there was a significant increase in the amount of NICD in the Hyp group. No significant difference was evident at P18. The Hyp HB-EGF group had no significant increase at P14.5 compared with the Nx HB-EGF group, and significantly less than the Hyp saline group ($n = 4$ mice for each group and age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **c**, Delta1 was increased at P14.5 only in the Hyp saline group (P11: $n = 4$

for each group; P14.5 and P18: $n = 5$ for each group and age; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). **d–g**, Representative $\times 40$ confocal images of subcortical white matter in the transgenic Notch reporter (TNR) mice, where eGFP is expressed upon activation of Notch effector C-promoter binding factor 1 (CBF1), a downstream transcriptional target of Notch. **h**, Histogram represents the number of eGFP⁺Olig2⁺ cells at P14.5 in the white matter. The Hyp saline group showed a significant increase in eGFP⁺Olig2⁺ cells, corresponding to enhanced Notch activation in oligodendrocyte lineage cells. This contributes to delayed maturation of oligodendrocyte lineage cells observed after hypoxia ($n = 4$ mice per group; one-way ANOVA, Bonferroni post-hoc test for individual comparisons). Histogram is presented as means \pm s.e.m. **d–g**, Scale bars, 50 μ m. ** $P < 0.05$; *** $P < 0.01$.

C/EBP α poises B cells for rapid reprogramming into induced pluripotent stem cells

Bruno Di Stefano^{1,2}, Jose Luis Sardina^{1,2*}, Chris van Oevelen^{1,2*}, Samuel Collombet^{3,4,5}, Eric M. Kallin^{1,2†}, Guillermo P. Vicent^{1,2}, Jun Lu⁶, Denis Thieffry^{3,4,5}, Miguel Beato^{1,2} & Thomas Graf^{1,2,7}

CCAAT/enhancer binding protein- α (C/EBP α) induces transdifferentiation of B cells into macrophages at high efficiencies and enhances reprogramming into induced pluripotent stem (iPS) cells when co-expressed with the transcription factors Oct4 (Pou5f1), Sox2, Klf4 and Myc (hereafter called OSKM)^{1,2}. However, how C/EBP α accomplishes these effects is unclear. Here we find that in mouse primary B cells transient C/EBP α expression followed by OSKM activation induces a 100-fold increase in iPS cell reprogramming efficiency, involving 95% of the population. During this conversion, pluripotency and epithelial–mesenchymal transition genes become markedly upregulated, and 60% of the cells express Oct4 within 2 days. C/EBP α acts as a ‘path-breaker’ as it transiently makes the chromatin of pluripotency genes more accessible to DNase I. C/EBP α also induces the expression of the dioxygenase Tet2 and promotes its translocation to the nucleus where it binds to regulatory regions of pluripotency genes that become demethylated after OSKM induction. In line with these findings, overexpression of Tet2 enhances OSKM-induced B-cell reprogramming. Because the enzyme is also required for efficient C/EBP α -induced immune cell conversion³, our data indicate that Tet2 provides a mechanistic link between iPS cell reprogramming and B-cell transdifferentiation. The rapid iPS reprogramming approach described here should help to fully elucidate the process and has potential clinical applications.

Reprogramming of somatic cells into induced pluripotent stem (iPS) cells by the OSKM transcription factors (also called Yamanaka factors) can be divided into a stochastic and a deterministic phase^{4,5}. During this process cells undergo a mesenchymal–epithelial transition (MET) and activate endogenous pluripotency genes, paralleled by changes in histone marks, nucleosome positioning and chromatin accessibility⁶. In addition, their promoters become de-methylated, preceded by oxidation of methylated CpGs through the dioxygenases Tet2 and Tet1, both of which have been implicated in the establishment of pluripotency^{7,8}. Earlier work showed that co-expression of C/EBP α with OSKM increases the reprogramming efficiency of B cells ~15-fold, reaching ~3% of the population¹. Here we describe that a pulse of C/EBP α followed by OSKM overexpression permits the rapid reprogramming of B cells into iPS cells by activating Tet2 and facilitating accessibility of pluripotency gene promoters to Oct4 binding. Highly efficient reprogramming of somatic cells has recently also been reported with a loss-of-function approach⁹.

Committed B-cell precursors (hereafter referred to as B cells) can be induced to transdifferentiate into macrophages at 100% efficiency by forced C/EBP α expression, deregulating ~7,500 genes¹⁰. Reasoning that chromatin of cells in transition might be more ‘open’ than that of end stages, we tested the effect of transiently exposing B cells to C/EBP α , followed by OSKM expression. B cells were isolated from the bone marrow of reprogrammable mice (containing a tetracycline-controlled transactivator (rtTA) and a doxycycline-responsive OSKM cassette¹¹),

infected them with C/EBP α -ER-hCD4 retrovirus, sorted human CD4⁺ cells 4 days later and incubated them for different times with β -estradiol (E2) followed by a wash-out (Fig. 1a). Subsequently, OSKM was induced by doxycycline treatment and Nanog⁺ colonies scored 12 days post induction (d.p.i.). B cells continuously co-expressing C/EBP α with OSKM showed an 11-fold enhancement in reprogramming efficiency compared to cells induced with OSKM alone (B+OSKM cells), confirming earlier reports^{1,12}. In contrast, cells treated for 18 h with E2 and then treated with doxycycline (B α ' + OSKM cells) exhibited a 103-fold colony increase, with cells pulsed for 6 h already showing a 74-fold increase (Fig. 1b–d). Clonal assays showed that 92–94% of viable colonies were Nanog⁺ after 12 days (Extended Data Fig. 1a–c). OSKM induction of B cells pre-treated with E2 did not increase Nanog⁺ colony numbers nor Oct4 expression levels (Extended Data Fig. 2a). A mutant of C/EBP α (BRM-2 (ref. 13)) defective for DNA binding failed to enhance iPS cell generation (Extended Data Fig. 2b). Induced B α ' + OSKM cells remained >50% viable (Extended Data Fig. 2c). C/EBP α pulses after OSKM induction had no effect on reprogramming efficiency (Extended Data Fig. 2d), indicating that C/EBP α acts as a path-breaker for OSKM-induced reprogramming. Stable iPS cell lines derived from B α ' + OSKM cells (α iPS cells) displayed similar gene expression profiles as embryonic stem (ES) cells, differentiated into all three germ layers *in vitro* and *in vivo*, and efficiently contributed to coat colour chimaerism (Fig. 1e and Extended Data Fig. 3a–e).

B α ' + OSKM-derived iPS cell colonies could be identified as early as 4 d.p.i. and their numbers increased modestly after 8 days (Fig. 1f, g), whereas B+OSKM colonies continued to increase 8–10 d.p.i. (Fig. 1f, g). Retroviral hCD4 expression was found to be silenced within 2–4 days in B α ' + OSKM cells compared to ~8 days in B+OSKM cells (Fig. 1h). To test the effect of C/EBP α on OSKM-induced transgene independence, B α ' and B cells were doxycycline-treated for different times (Fig. 1i). The first transgene-independent iPS cell colonies were observed after 4 days for B α ' + OSKM cells, compared to 9–10 days for B+OSKM controls (Fig. 1j). Together these results show that the C/EBP α pulse accelerates iPS cell reprogramming by 4–6 days.

Gene expression analyses of B α ' + OSKM cells 8 d.p.i. showed the upregulation of 764 out of 1,668 genes expressed more highly in ES cells than in B cells, including all well described pluripotency genes (Fig. 2a). Unsupervised hierarchical clustering analysis revealed that 8 d.p.i. B α ' + OSKM cells clustered with ES/iPS cells (Extended Data Fig. 4a). A large part of known pluripotency genes were activated 2–6 d.p.i. (Fig. 2b, c and Extended Data Fig. 4b), reaching levels comparable to α iPS cells and ES cells within ~1 week (Fig. 2c and Extended Data Fig. 4c). In contrast, they remained essentially silent in B+OSKM cells (Fig. 2b, c). In addition, five genes described to be activated very late (21 d.p.i.) during fibroblast reprogramming¹⁴ became upregulated within 2–4 days in B α ' + OSKM cells (Extended Data Fig. 4d). Using Oct4–GFP reporter mice¹² crossed with reprogrammable mice¹¹, B α ' + OSKM

¹Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003 Barcelona, Spain. ²Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain. ³Ecole Normale Supérieure, Institut de Biologie de l'ENS, 45 Rue d'Ulm, Paris F-75005, France. ⁴Inserm, U1024, Paris F-75005, France. ⁵CNRS, UMR 8197, Paris F-75005, France. ⁶Yale Cancer Center, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg Lluís Companys 23, 08010 Barcelona, Spain. [†]Present address: Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York 10065, USA.

*These authors contributed equally to this work.

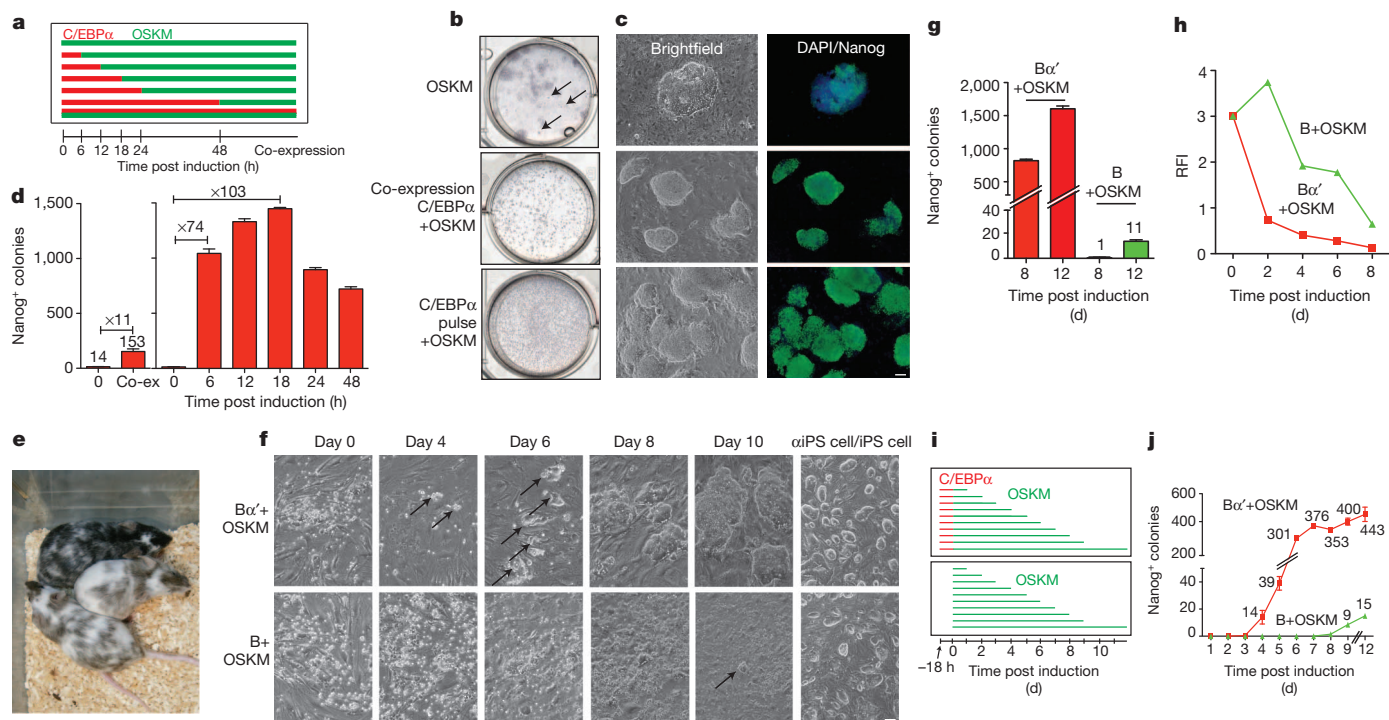


Figure 1 | Effect of transient C/EBP α expression on OSKM-induced iPS cell reprogramming of B cells. **a**, Experimental strategy. **b**, Six-well plates with alkaline-phosphatase-positive iPS cell colonies at 12 d.p.i. obtained from B-cell cultures treated with doxycycline to induce OSKM only (B + OSKM), cells co-expressing C/EBP α with OSKM, or cells treated for 18 h with E2 to activate C/EBP α followed by OSKM induction (B α' + OSKM). **c**, Nanog $^{+}$ iPS cell colonies at 12 d.p.i. (scale bar, 100 μ m). **d**, Quantification of 12 d.p.i. Nanog $^{+}$ colonies after OSKM induction in B cells pre-treated with C/EBP α for the indicated pulse lengths. Error bars indicate s.d. ($n = 3$). **e**, Chimeric mice

obtained after blastocyst injection of α iPS cell clone 12. **f**, Doxycycline-induced B α' + OSKM cultures with arrows indicating earliest detectable ES-cell-like colonies (scale bar, 100 μ m). **g**, Nanog $^{+}$ colonies of B α' + OSKM and B + OSKM cultures at 8 and 12 d.p.i. Error bars indicate s.d. ($n = 3$). **h**, Retroviral silencing after induction of B α' + OSKM and B + OSKM cells measured by hCD4 expression. RFI, relative fluorescence intensity. **i**, **j**, Experimental strategy to determine OSKM transgene independence (**i**) and formation of Nanog $^{+}$ iPS cell colonies (**j**) at 12 d.p.i. Error bars indicate s.d. ($n = 3$).

cells became 30% GFP $^{+}$ at 4 d.p.i. and 60% GFP $^{+}$ at 8 d.p.i., whereas B + OSKM cells remained negative (Extended Data Fig. 4e). Serum-free conditions further accelerated the process with 62% of the B α' + OSKM cells becoming GFP $^{+}$ already at 2 d.p.i. and 95% at 4 d.p.i. (Fig. 2d, e), with a similar viability as in serum containing medium. Induced B + OSKM cells died without serum.

We next scored the expression of MET genes^{15,16}. B α' + OSKM cells upregulated all epithelial genes tested and expressed E-cadherin 2–4 d.p.i., again with B + OSKM cells remaining negative (Extended Data Fig. 5a, b). In line with the possibility that the C/EBP α pulse is capable of initiating an epithelial–mesenchymal transition (EMT)¹⁷, mesenchymal genes encoding TGF- β pathway members, transcription factors and collagens were first upregulated by the C/EBP α pulse and then switched off after OSKM induction (Extended Data Fig. 5a, e). To compare gene expression changes during OSKM-induced reprogramming of B α' cells with C/EBP α -induced transdifferentiation of B cells we selected genes selectively expressed in macrophages (263) and B cells (83) (Supplementary Table 1). Most of these genes were already up- or downregulated in B α' cells, respectively, as expected from the induced conversion (Extended Data Fig. 6a). After OSKM induction, B-cell genes became further silenced whereas macrophage genes elicited a more heterogeneous response (Extended Data Fig. 6b, c). B α' + OSKM cells at 2–4 d.p.i. co-expressed E-cadherin and the macrophage markers CSF-1R and Mac1, whereas B + OSKM cells remained negative and α iPS cells expressed exclusively E-cadherin (Extended Data Fig. 6d). Sorted Mac1 $^{+}$ E-cadherin $^{+}$ cells and Mac1 $^{-}$ E-cadherin $^{+}$ cells yielded similar proportions of iPS cell colonies, indicating that expression of Mac1 has no role in iPS cell reprogramming (Extended Data Fig. 6e). Analysis of the 18 h pulsed B cells by RNA-seq revealed 1,418 upregulated and 552 downregulated genes (Extended Data Fig. 7a). Gene set enrichment analysis (GSEA)

showed a strong correlation between genes downregulated in B α' cells with sets of B-cell-specific genes, and upregulated genes with genes related to EMT, collagen expression and the TGF- β pathway (Extended Data Fig. 7b), confirming the array data. C/EBP α chromatin immunoprecipitation sequencing (ChIP-seq) analysis revealed 1,766 DNA regions bound by the protein in B α' cells of which \sim 20% were within 50 kb of genes whose expression changed (Extended Data Fig. 7c–g). These included two regulatory sites in *Tet2* (Fig. 3a, ref. 3) as well as sites close to lymphoid, MET and TGF- β pathway genes. *Tet2* became upregulated \sim 1.6-fold after the 18-h pulse and was further activated by OSKM, reaching levels comparable to ES cells within 2 days (Fig. 3b, c). In contrast, mouse embryonic fibroblasts (MEFs) pulsed with C/EBP α did not upregulate the gene (Extended Data Fig. 8a). Tet2 overexpression enhanced iPS cell reprogramming of B cells \sim 3-fold (Fig. 3d), supporting the reported role of Tet2 in iPS cell formation^{7,8}. Because the Tet2-induced 5-hydroxymethylcytosine (5hmC) modification has been described to participate in the maintenance of pluripotency, we determined 5hmC levels by hydroxymethylated DNA immunoprecipitation followed by qPCR (hMeDIP-qPCR) in B α' cells at regulatory regions of nine pluripotency genes for which hydroxymethylation levels change during reprogramming¹⁸. 5hmC levels were significantly increased for all genes tested (Fig. 3e and Extended Data Fig. 8b) and Tet2 binding was detected at the same regions (Fig. 4f and Extended Data Fig. 8c). Unexpectedly, Tet2 was predominantly cytoplasmic in uninduced B cells and nuclear after the C/EBP α pulse (Fig. 3g). The nuclear localization was maintained after sustained expression of C/EBP α (72 h) (Extended Data Fig. 8d). No protein translocation was observed in B cells treated with E2 alone or infected with a control vector (Extended Data Fig. 8e) and MEFs exhibited predominantly nuclear Tet2 localization (Extended Data Fig. 8f). Because hydroxymethylation is required

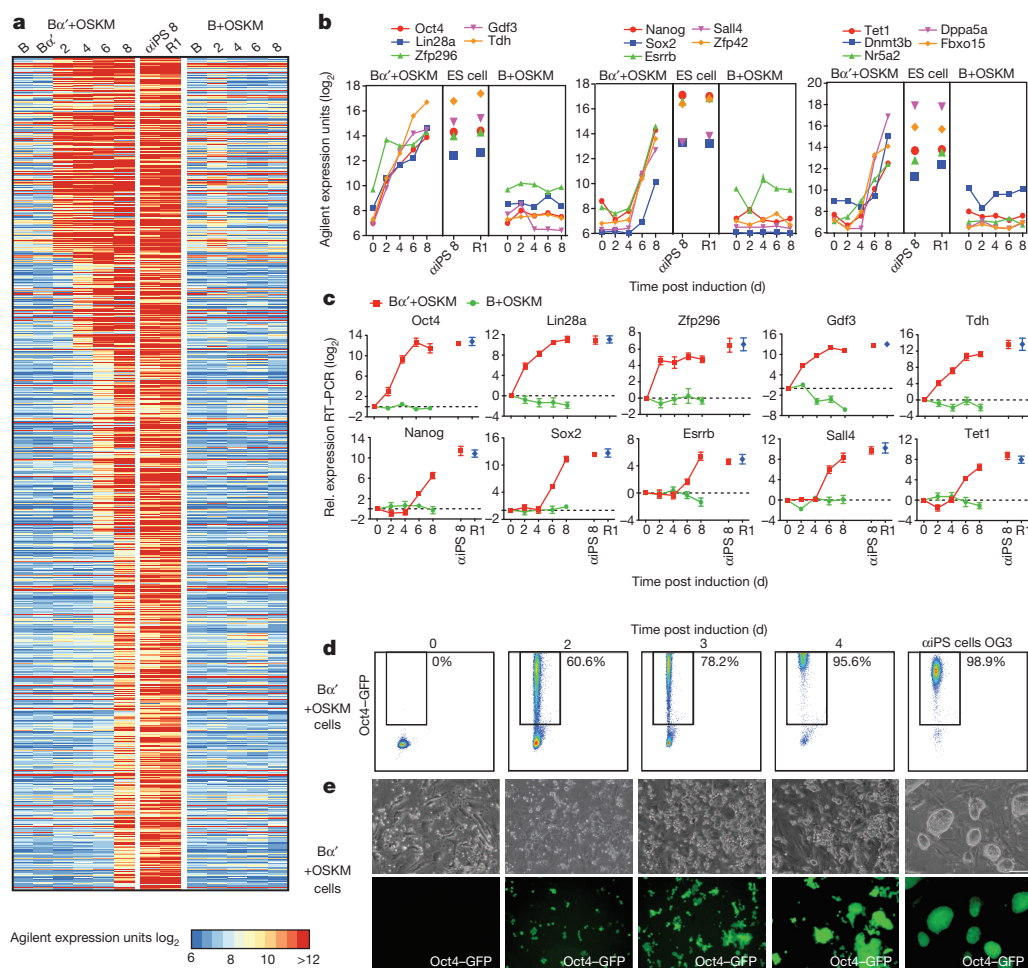


Figure 2 | Upregulation of pluripotency genes. **a**, Heat map of Agilent expression arrays showing upregulation of genes expressed >4-fold higher in ES cells than in B cells at the indicated time points after doxycycline treatment of B α' +OSKM cells and B+OSKM cells, with α iPS cells (clone 8) and ES cells (R1) as controls. **b**, Array expression kinetics of three groups of pluripotency

genes from B α' +OSKM cells, α iPS cells and ES cells and B+OSKM cells. **c**, qRT-PCR validation of selected genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, **e**, Oct4-GFP expression analysed by FACS and microscopy in B α' +OSKM cells (scale bar, 100 μ m).

for active DNA demethylation, we analysed the DNA methylation status of the *Oct4* and *Nanog* promoters using bisulphite sequencing (Fig. 3h). Whereas the *Oct4* promoter became partially demethylated in B α' +OSKM as early as 2 d.p.i. (Fig. 3h), the *Nanog* promoter only became demethylated at 8 d.p.i., consistent with the observed delayed upregulation of the gene (Fig. 2b, c).

The 5hmC changes at pluripotency genes raised the possibility that the C/EBP α pulse eliminates epigenetic barriers, increasing chromatin accessibility to the OSKM transcription factors. To test this, we performed quantitative DNase I sensitivity assays on selected regulatory regions of pluripotency genes. This revealed increased accessibility to the enzyme of all sites tested in the 18-h pulsed B cells whereas control regions and pluripotency sites in control B+OSKM cells showed no difference (Fig. 4a). Interestingly, after prolonged induction of C/EBP α (72 h) the chromatin of pluripotency genes returned to the closed state as seen in untreated B cells (Extended Data Fig. 9a). In addition, no demethylation of the *Nanog* and *Oct4* promoters was observed (Extended Data Fig. 9b). We next tested Oct4 occupancy in OSKM-induced B α' cells at sites within enhancers and promoters of pluripotency genes bound by Oct4 in ES cells¹⁹ (Fig. 4b). At 2 d.p.i. all of these sites showed significant Oct4 occupancy, whereas they remained unbound in B+OSKM cells (Fig. 4b). Together, these findings show that the C/EBP α pulse transiently induces chromatin remodelling at regulatory regions of pluripotency genes, making them rapidly accessible to Oct4 binding.

Our results show that C/EBP α poises B cells for highly efficient and almost immediate reprogramming into iPS cells by OSKM transcription factors, with MET and pluripotency network activation occurring concomitantly (summarized in Extended Data Fig. 9c). In addition, our data indicate that the C/EBP α pulse already initiates an EMT, reminiscent of the recently described EMT-MET transition during reprogramming of fibroblasts¹⁷. The single cell cloning and Oct4-GFP reporter experiments suggest that essentially 100% of the poised cells can be reprogrammed into iPS cells. Several lines of evidence suggest that the rapid demethylation of the *Oct4* promoter involves Tet2: the C/EBP α pulse directly binds to and upregulates the *Tet2* gene and shuttles the protein into the nucleus. Then Tet2 binds to regulatory sites of *Oct4* and other pluripotency genes, converting 5mC residues into 5hmC (summarized in Fig. 4c). This modification might directly cause the gene's de-repression or be linked to demethylation and/or induction of chromatin remodelling⁸. The effect of C/EBP α is highly specific as only the closely related factor C/EBP β , but not the lineage-instructive factors GATA1 (ref. 20), nor Mash1 or MyoD expressed in fibroblasts, was active (Extended Data Fig. 10a–i). C/EBP α also sensitized T-cell precursors²¹ but not fibroblasts (Extended Data Fig. 10j, k). However, when C/EBP α was co-expressed with PU.1, a condition that mediates transdifferentiation to macrophage-like cells²², it induced a ~6-fold enhancement of iPS cell reprogramming, whereas PU.1 alone had no effect (Extended Data Fig. 10k). This indicates that a myeloid determinant sensitizes somatic cells for iPS cell reprogramming, in line with

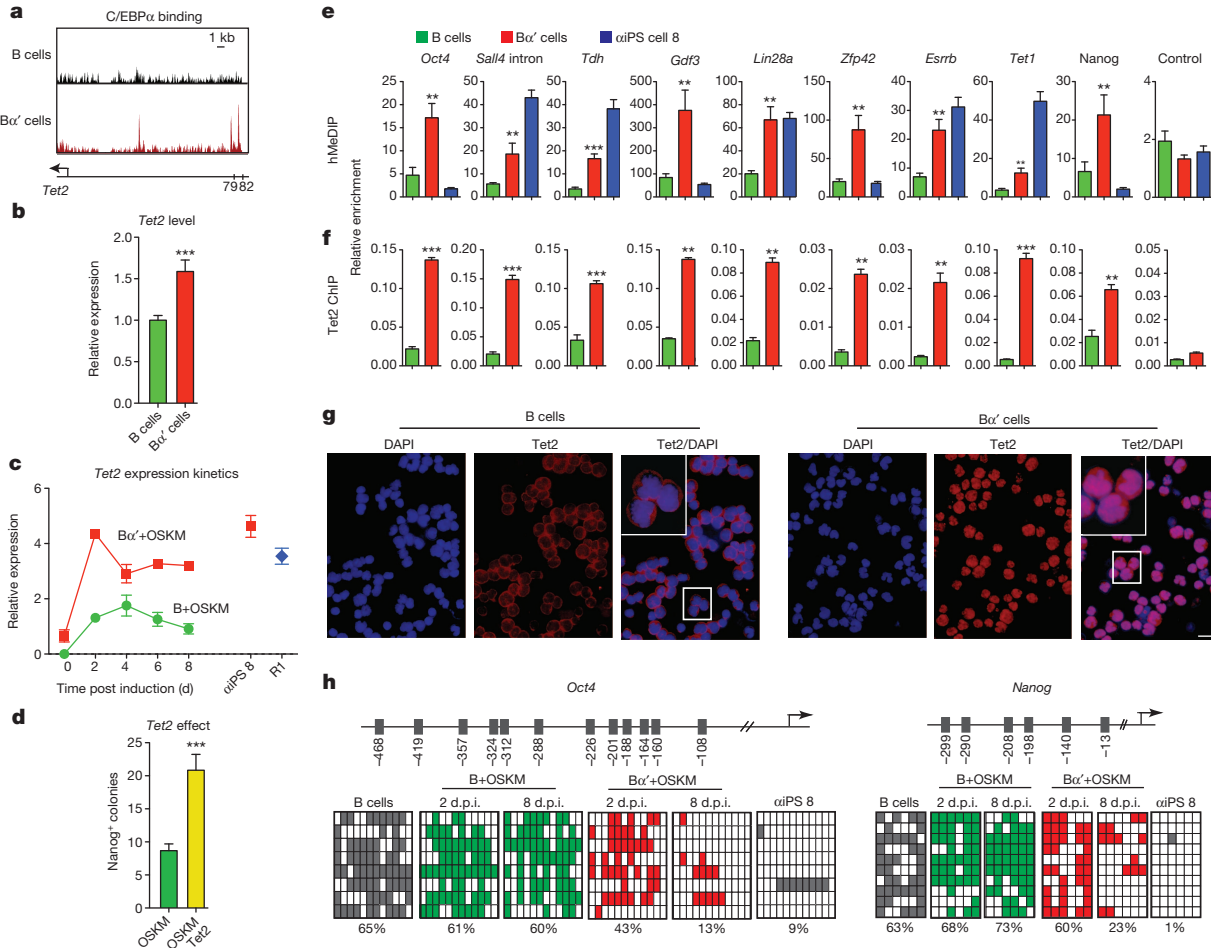


Figure 3 | Tet2 activation, cytosine hydroxymethylation and DNA methylation. **a**, C/EBP α binding at two enhancers upstream of the *Tet2* gene. **b**, c, qRT-PCR for *Tet2* expression. Error bars indicate s.d. ($n = 6$). Student's *t*-test *** $P < 0.001$ relative to control. **d**, *Tet2* overexpression effect on B-cell reprogramming. Error bars indicate s.d. ($n = 6$). Student's *t*-test *** $P < 0.001$ relative to control. **e**, hMeDIP followed by qPCR at regulatory regions of selected pluripotency genes. Error bars indicate s.d. ($n = 3$). Student's *t*-test

* $P < 0.01$, *** $P < 0.001$ relative to B cells. **f**, Tet2 binding sites determined by ChIP of the regions in **e**. Error bars indicate s.d. ($n = 3$). Student's *t*-test *** $P < 0.01$, *** $P < 0.001$ relative to B cells. **g**, Immunofluorescence of Tet2 protein in B cells and Bx₆ cells (scale bar, 30 μ m). **h**, CpGs analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters, with their position indicated on the top. Filled rectangles represent methylated CpGs; empty rectangles represent unmethylated residues.

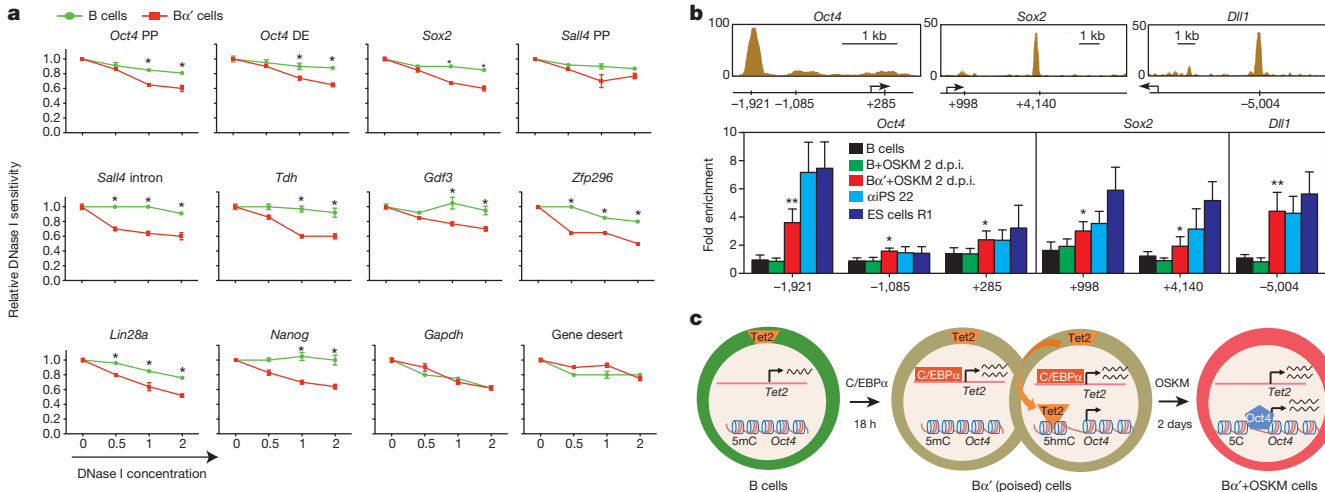


Figure 4 | Chromatin accessibility and summary schemes. **a**, DNase I sensitivity at regulatory regions in B cells and Bx' cells determined by enzyme titration. Error bars indicate s.d. ($n = 3$). Student's *t*-test * $P < 0.05$ relative to control. DE, distal enhancer; PP, proximal promoter. **b**, Oct4 binding sites

examined at the *Pou5f1*, *Sox2* and *Dll1* loci, with Oct4 peaks corresponding to ES cells¹⁹. Oct4 binding by ChIP-qPCR analysis. Student's *t*-test **P* < 0.05, ***P* < 0.01 relative to control. Error bars indicate s.d. (*n* = 3). **c**, Model of reprogramming mechanism involving Tet2.

the finding that macrophage/granulocyte progenitors, the formation of which is driven by C/EBP α ²³, show the highest described reprogramming efficiency of somatic cells (25%)¹². The C/EBP α effect might recapitulate an embryonic function of the gene as it is required, in combination with C/EBP β , for the formation of functional trophoblasts²⁴, cells with intriguing similarities to macrophages²⁵.

METHODS SUMMARY

Somatic cells used for iPS cell reprogramming were derived from the reprogrammable mouse¹¹. ES cells and iPS cells were cultured on mitomycin-C-treated MEFs in ES cell medium containing 15% FBS and 1,000 U ml⁻¹ of LIF. For B- and T-cell reprogramming experiments the medium was supplemented with cell-type-specific cytokines¹². OSKM expression was induced with 2 ng ml⁻¹ of doxycycline. α iPS cells and iPS cells correspond to individual iPS cell clones obtained after expansion into stable lines of OSKM induced B α' and B cells, respectively. The total number of iPS cell colonies was scored after alkaline phosphatase or Nanog staining. For qRT-PCR and microarray analyses, B cells at different time points during reprogramming, as well as ES cells and iPS cells, were trypsinized and FACS sorted to remove feeder cells and dead cells. Total RNA was extracted using the miRNeasy kit (Qiagen). Microarrays were done using Agilent 8X60K expression arrays. Details on injections into blastocysts to test for chimaerism are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 April; accepted 15 November 2013.

Published online 15 December 2013.

- Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Kallin, E. M. *et al.* Tet2 facilitates the derepression of myeloid target genes during CEBP α -induced transdifferentiation of pre-B cells. *Mol. Cell* **48**, 266–276 (2012).
- Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
- Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
- Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835–850 (2011).
- Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374 (2013).
- Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
- Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
- Di Tullio, A. *et al.* CCAAT/enhancer binding protein α (C/EBP α)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108**, 17016–17021 (2011).
- Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods* **7**, 56–59 (2010).
- Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41**, 968–976 (2009).
- Porse, B. T. *et al.* E2F repression by C/EBP α is required for adipogenesis and granulopoiesis *in vivo*. *Cell* **107**, 247–258 (2001).
- Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012).
- Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
- Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
- Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nature Cell Biol.* **15**, 829–838 (2013).
- Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
- Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
- Kulesa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thrombocytes, and erythroblasts. *Genes Dev.* **9**, 1250–1262 (1995).
- Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP α and PU.1 transcription factors. *Immunity* **25**, 731–744 (2006).
- Feng, R. *et al.* PU.1 and C/EBP α / β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
- Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α -deficient mice. *Proc. Natl Acad. Sci. USA* **94**, 569–574 (1997).
- Begay, V., Smink, J. & Leutz, A. Essential requirement of CCAAT/enhancer binding proteins in embryogenesis. *Mol. Cell. Biol.* **24**, 9744–9751 (2004).
- Guilbert, L., Robertson, S. A. & Wegmann, T. G. The trophoblast as an integral component of a macrophage-cytokine network. *Immunol. Cell Biol.* **71**, 49–57 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Batlle for generating the chimaeric mice, M. Stadtfeld, D. Egli, P. Cosma, G. Filion, B. Lehner and J. Valcarcel for critical reading of the manuscript, V. Broccoli, P. Muñoz and S. Aznar Benitah for constructs and antibodies and H. Schoeler for the Oct4 reporter mice. This work was supported by the Ministerio de Educacion y Ciencia, SAF.2007-63058 and AGAUR 2009 SGR768. B.D.S. holds a La Caixa International PhD Fellowship.

Author Contributions T.G. and B.D.S. conceived the idea for this project, designed the experiments and wrote the paper. B.D.S. performed the cell culture, animal experiments and molecular biology and B.D.S. and J.L.S. the methylation analyses. J.L.S. performed the studies on Tet2 and J.L.S. and C.v.O. the ChIP experiments. S.C., E.M.K. and D.T. were responsible for the bioinformatics and G.P.V. and M.B. for the DNase I experiments and analyses. J.L. provided the Tet2 overexpressing plasmids.

Author Information The array data are available from the Gene Expression Omnibus under the accession number GSE52397. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.G. (Thomas.Graf@crg.eu).

METHODS

Mice. The reprogrammable mouse containing a tetracyclin-inducible OSKM cassette as well as rTA has been described previously¹¹.

Cell cultures. ES cells and iPS cells were cultured on mitomycin-C-treated MEF feeder cells in KO-DMEM (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, non-essential amino acids, β -mercaptoethanol, 1,000 U ml⁻¹ LIF (ES cell medium) and 15% fetal bovine serum (FBS, Invitrogen).

MEF cultures were established by trypsin digestion of mouse embryos (embryonic day 13.5) and the resulting cells cultured in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin. CD19⁺ pro-B and pre-B cells were isolated from bone marrow using monoclonal antibodies to CD19 (1D3), obtained from BD Pharmingen, using MACS (Miltenyi Biotech). Double-negative (DN) T cells were prepared from thymii of 4-week-old mice by lineage depletion with Streptavidin microbeads. The following biotin-conjugated antibodies were used to label lineage (lin)-positive thymocytes before lineage depletion: CD3 (145-2C11), CD4 (GK1.5), CD8 (53-6.7), B220 (RA3-6B2), CD19 (1D3), Mac1/CD11b (M1/70), Gr1 (RB6-8C5), Ter119 (TER-119), I-A/I-E (2G9). The purity of the sorted cell fractions was confirmed by FACS using an LSR II machine (BD). After isolation, B cells were expanded in RPMI medium supplemented with 10% FBS and IL-7 for 1 week before inducing reprogramming. T cells were grown RPMI medium supplemented with 10% FBS supplemented with IL-7, Flt3 and SCF.

Reprogramming. Reprogramming experiments were conducted in gelatinized plates seeded with a feeder layer of the OP9 stromal cell line, using ES cell medium supplemented with 2 μ g ml⁻¹ of doxycycline and 15% FBS. For the reprogramming of B cells, IL-4 (10 ng ml⁻¹), IL-7 (10 ng ml⁻¹) and IL-15 (2 ng ml⁻¹) were added to the medium (ref. 13). B cells were seeded at a density of 500 cells cm⁻² in 6-well plates. For the conditioning of B α' cells, C/EBP α -infected sorted B cells were exposed for 18 h to 100 nM of β -estradiol (E2) followed by inducer wash-out and addition of ES cell medium supplemented with doxycycline and cytokines. Serum-free iPS reprogramming was performed in serum-free ES cell medium (KO-DMEM supplemented with 15% knockout serum replacement, L-glutamine, penicillin/streptomycin, non-essential amino acids, β -ME, 1,000 U ml⁻¹ LIF, N2 (100X), B27 (100X))⁹ supplemented with 2 μ g ml⁻¹ doxycycline to activate OSKM factors. 21 inhibitors were added to the medium 2 days after doxycycline induction.

For expansion of iPS cell lines, colonies with ES cell morphology were picked after doxycycline withdrawal at 12 d.p.i. OSKM MEFs were seeded on mitomycin-C-treated MEF feeders in ES cell medium containing FBS and induced with doxycycline.

Vectors and virus production and infection. C/EBP α -ER-GFP, C/EBP α -ER-CD4, C/EBP β , PU.1, C/EBP α -BRM-2 retroviruses and shPax5 lentivirus have been described previously^{1,2,13}. *Gata1* cDNA was cloned by PCR in the BglII/XhoI restriction sites of the MIG vector. The MyoD-ER plasmid was obtained from P. Muñoz, Mash1 from V. Broccoli. The retroviral vector for murine Tet2 overexpression was derived from MSCV-based pMIRWAY-puro-TET2²⁶ by replacing puro with EGFP. Virus production has been described previously²⁷. Briefly, HEK293T cells were co-transfected with vector plasmid and packaging plasmids using calcium phosphate transfection. Viral supernatants were collected 48–72 h later and concentrated by ultracentrifugation at 20,000g for 2 h at 20 °C. Viral concentrates were re-suspended in 1 \times PBS and stored at –80 °C. Infection of MEFs was carried out in medium containing 5 μ g ml⁻¹ polybrene, and infection of B cells by centrifugation with concentrated virus for 2 h at 32 °C at 1000g in B-cell medium.

Alkaline phosphatase staining. For alkaline phosphatase staining, the cells were fixed in 2% paraformaldehyde and then incubated for 15 min in NTMT solution (NaCl 100 mM, Tris A 1 M, Tris B 1 M, MgCl₂ 50 mM, Tween 0.1%) supplemented with BCIP e NBT (Roche).

Immunofluorescence assays. The cells were fixed with 4% paraformaldehyde, blocked and incubated with primary antibodies overnight at 4 °C. On the next day, the cells were exposed to secondary antibodies (all Alexa Fluor from Invitrogen) at room temperature for 1 h. The primary antibodies used were Nanog (Calbiochem), Oct4 (SC-5279), SMA (obtained from P. Muñoz), Tuj1 (obtained from V. Broccoli), Foxa2 (obtained from S. Aznar Benitah) and Tet2 (SC-61). Nuclear staining was performed with DAPI (Invitrogen).

Differentiation of iPS cells. Embryoid bodies (EBs) were derived by plating iPS cells at a concentration of 1.3 \times 10⁶ cells ml⁻¹ in non-adherent dishes in ES medium without LIF. After 4 days in suspension, cell aggregates were plated on gelatine-coated dishes in DMEM medium supplemented with 10% FBS for 10 days.

Teratoma assay. One million iPS cells were injected subcutaneously into SCID BEIGE mice. Three weeks after the injection, the tumours were surgically removed and embedded in paraffin. Teratomas were sectioned into 10- μ m-thick slices using a cryostat (Leica, CM1850 UV) and sections were stained with haematoxylin and eosin.

Chimaeric mice. For the chimaera formation assay, 10–15 iPS cells (C57BL/6J background, black coat colour) were injected into a 3.5 days post coitum (d.p.c.) blastocyst of CD1 mice (white coat colour) and transferred into pseudo-pregnant

2.5 d.p.c. CD-1 recipients for chimaera generation. Chimaerism was ascertained after birth by the appearance of black coat colour (from the C57BL/6J iPS cells) over the white background from the host pups.

FACS. Cell suspensions were stained with various antibodies from either BD Pharmingen or eBiosciences. Cells were analysed with an LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar). Primary antibodies used were CD19, Mac1, B220, Cdh1, CD115 (all from BD Biosciences), hCD4 (eBioscience) and Nanog (Calbiochem).

RNA isolation and quantification. RNA isolation of MEFs and B cells was done with the miRNeasy mini kit (Qiagen). RNA was collected from sorted live cells at 0, 2, 4, 6 and 8 d.p.i. To remove the feeders, ES cells and iPS cell clones at passage 2 or higher were seeded on gelatinized plates and processed with the above kits. RNA was eluted from the columns using RNase-free water or TE buffer and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystem).

Gene expression arrays and qRT-PCR. RNA samples (with an RNA integrity number (RIN) greater than 9) were subjected to transcriptional analyses using Agilent expression arrays. For hybridization, 500 ng of total RNAs were labelled using Agilent's QuickAmp labelling kit following manufacturer instructions and analysed using Agilent 8X60K expression arrays. For the expression data with multiple probes annotated to the same gene only the most dynamic probe as defined by all array hybridizations was considered. Genes with expression changes during reprogramming were defined as showing >4-fold differences between ES cells and uninduced B cells, and having a Student's *t*-test *P* value <0.05 between technical duplicates. B-cell- and macrophage-specific genes were defined by combining array data from a previous study²⁸. Lineage-specific genes were defined as >4-fold upregulated in one lineage over all of the others. Macrophage-specific genes were selected on the basis of >4-fold upregulation in primary macrophages and 5-day C/EBP α -induced macrophages as compared to B cells.

Unsupervised clustering was performed on Pearson correlation coefficients calculated between the indicated array samples using all genes that changed expression at least 2-fold between B cells and ES cells (*n* = 10,982). Data were analysed and visualized using R v2.15.1.

The array data are available from the Gene Expression Omnibus under the accession number GSE52397.

qRT-PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystem) and primers as listed in Supplementary Table 2. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 58 °C and 30 s at 72 °C.

Chromatin immunoprecipitation. ChIP experiments were performed as described previously²⁹. Antibodies against Oct4 (SC-8626) and C/EBP α (SC-61) were purchased from Santa Cruz Biotechnologies. Tet2 ChIP was performed as in ref. 3. Data were obtained from three biological replicates, three independent immunoprecipitations and three technical triplicates.

For ChIP-seq analysis, reads were mapped onto mouse mm9 genome using bowtie2 (v2.1.0) (parameter: –very-sensitive). Reads filtering was done using SAMtools to keep reads that map only once, with a quality score of 10 or more, and to remove duplicates. Peaks were called using HOMER (v4.3) (parameter: style factor) and peaks were selected to control the false discovery rate at 0.001. Peaks annotation and genes association (Extended Data Fig. 7b–d) was performed using HOMER and BedTools (v 2.17).

Circular plot (Extended Data Fig. 7e) was done using Circos.

The sequencing data are available under the Gene Expression Omnibus number GSE52397.

DNA methylation. DNA was extracted using the Blood & Cell Culture DNA Mini kit (Qiagen). Bisulphite treatment of DNA was achieved using the EpiTect bisulphite kit (Qiagen) according to the manufacturer's instructions. The resulting modified DNA was amplified by polymerase chain reaction (PCR) using primer listed in Supplementary Table 2. The resulting amplified products were gel-purified (Qiagen), subcloned into the pGEM-T Vector Systems (Promega), and sequenced using the T7 and SP6 primers.

DNase I assay. Chromatin samples obtained as described before from two biological replicates were subjected to DNase I digestion. Briefly, 2 μ g of chromatin were treated with 0.5, 1, and 2 units of DNase I (Roche) for 3 min at 37 °C in 1 \times DNase incubation buffer. Control samples were incubated in the absence of DNase I. Reactions were terminated by adding EDTA (40 mM final concentration) and the crosslinking was reversed by incubating the samples at 65 °C. After 6 h, proteinase K (40 μ g ml⁻¹ final concentration) was added to each reaction and incubated overnight at 37 °C. After careful phenol-chloroform extractions, the DNA was quantified and used as template for Real Time-PCR reactions using specific primers.

hMeDIP. DNA extraction and IP (DIP) was carried out by a previously published protocol (<http://www.epigenome-noe.net/>, PROT33) with the following alterations.

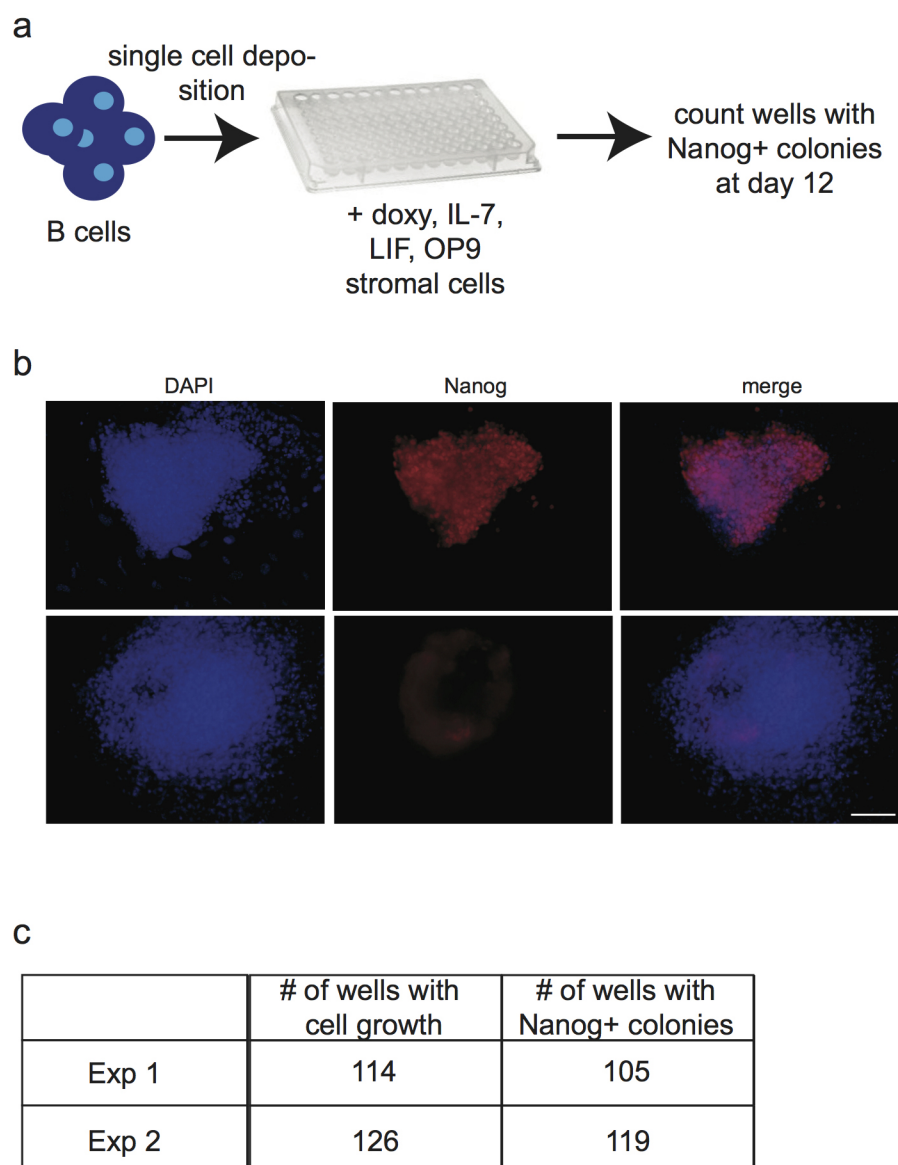
Genomic DNA was fragmented to a mean size of 350 bp using a Bioruptor (Diagenode) for 7×30 s cycles. DNA (re-suspended in 1× IP buffer) was incubated with 5 µg of anti-OHMeC antibody (Active Motif) at 4 °C for 2 h, and 40 µl of a 1:1 mixture of BSA-blocked Protein A (Millipore 16-125) and Protein G (Millipore 16-266) agarose beads were added, followed by an additional 2 h incubation at 4 °C. After washing 3 times in 1× IP buffer, immune complexes were released by incubation with 70 µg Proteinase K, and DNA was extracted once with phenol, extracted once with chloroform, and recovered by EtOH precipitation for qPCR analysis. The 5hmC enrichment in the samples was assessed by qPCR and the values were normalized against *Gapdh* promoter.

RNA-seq. RNA-seq was performed as described previously³⁰. For RNA-seq analysis, reads were aligned onto mouse mm9 genome using STAR and Refseq mm9 annotation for splicing. Reads were filtered to keep only uniquely mapped reads with a maximum of two mismatches. Gene read count was performed using HTseq-count (parameter: mode = union, stranded, features = exons, attribute = gene_id) with Refseq mm9 annotation, and data were normalized using Bioconductor DESeq library (parameter: method = blind, sharingMode = fit-only, fitType = local). Genes with a read count above 300 in all conditions (set as a minimal threshold

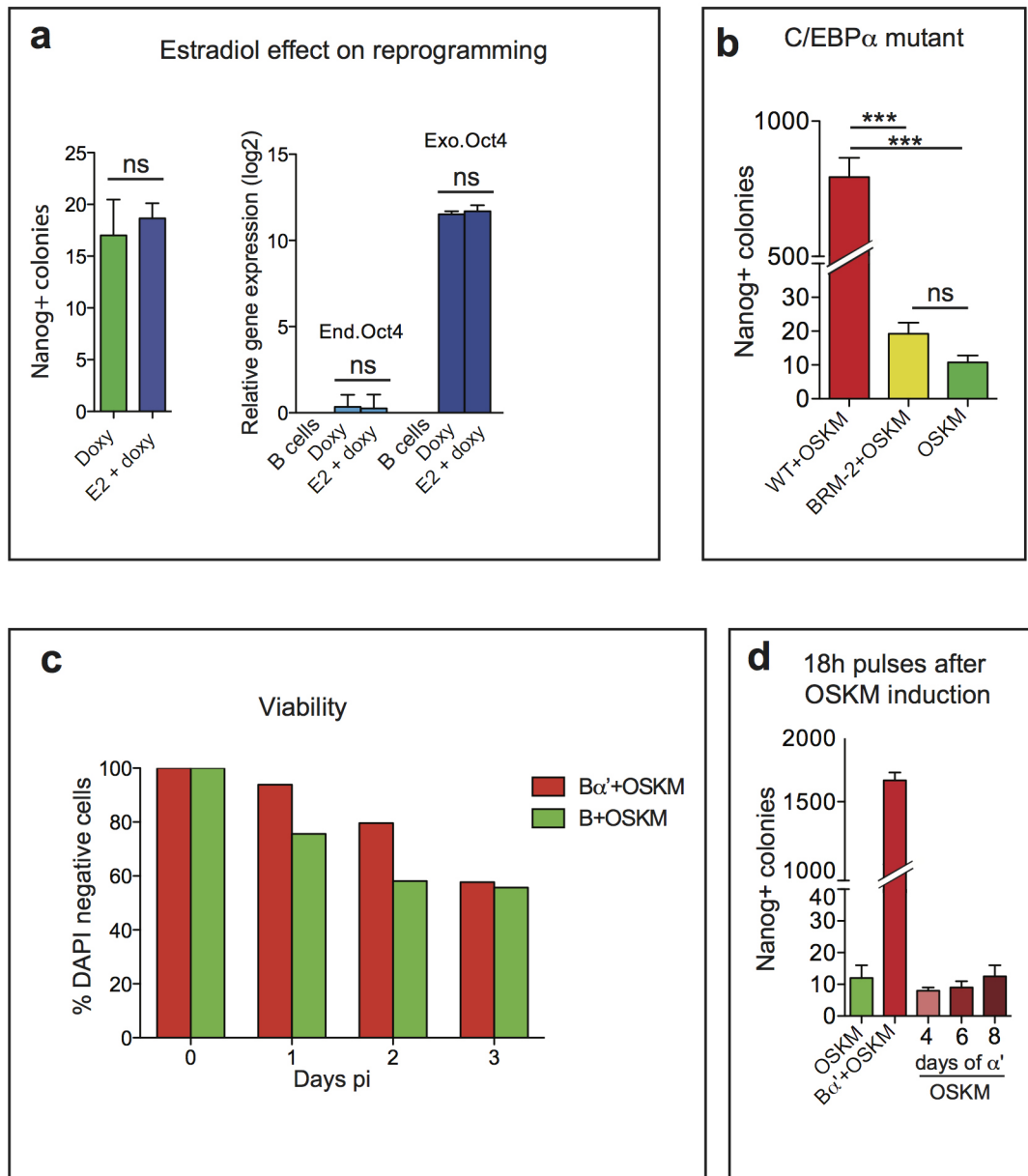
for functional expression) and genes with a 1.5-fold difference in normalized reads count between B cells and B α ' cells were considered as differentially expressed. The sequencing data are available under the Gene Expression Omnibus number GSE52397.

Statistical analysis and replicates. All data presented are representative of at least three independent experiments that yielded similar results. Statistical analyses were performed using Prism software (GraphPad).

26. Cheng, J. *et al.* An extensive network of TET2-targeting microRNAs regulates malignant hematopoiesis. *Cell Rep.* **5**, 471–481 (2013).
27. Di Stefano, B. *et al.* A microRNA-based system for selecting and maintaining the pluripotent state in human induced pluripotent stem cells. *Stem Cells* **29**, 1684–1695 (2011).
28. Di Tullio, A. & Graf, T. C/EBP α bypasses cell cycle-dependency during immune cell transdifferentiation. *Cell Cycle* **11**, 2739–2746 (2012).
29. van Oevelen, C. *et al.* A role for mammalian Sin3 in permanent gene silencing. *Mol. Cell* **32**, 359–370 (2008).
30. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).

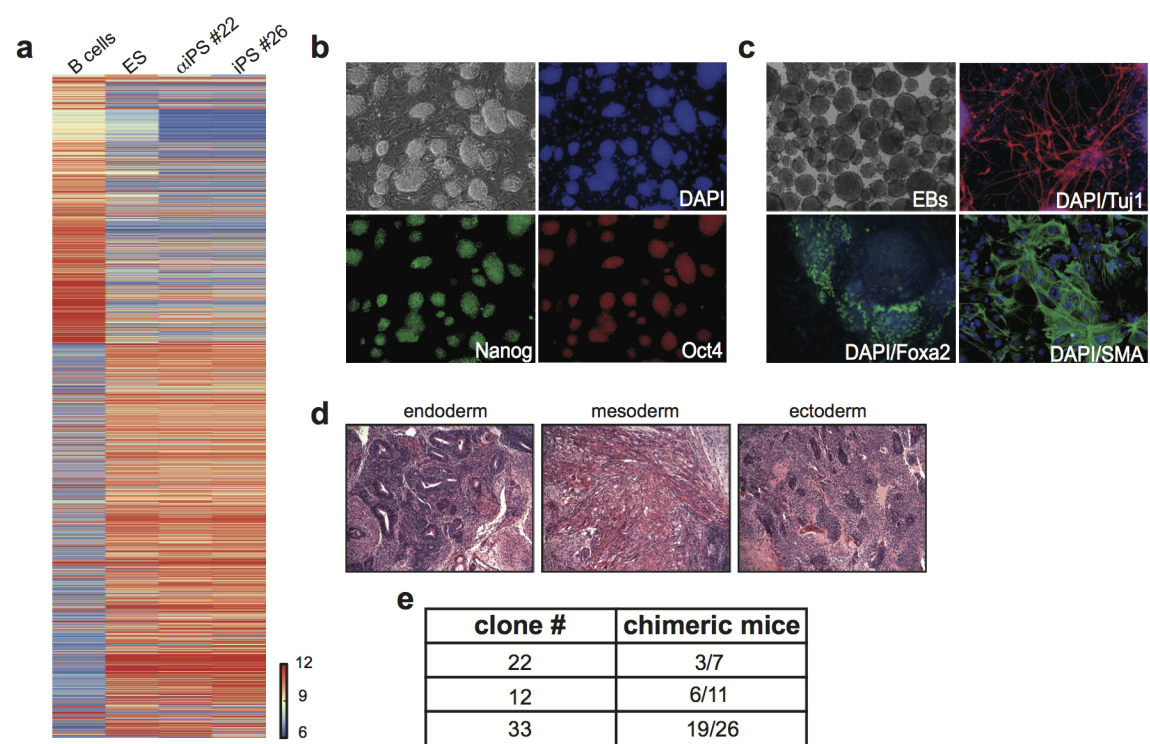


Extended Data Figure 1 | Single cell reprogramming. **a**, $B\alpha'$ cells were plated as single cells in 96-well plates on OP9 feeders in medium conditioned with IL-7, IL-15 and IL-4 cytokines. **b**, Representative Nanog-positive (upper panels) and Nanog-negative (lower panels) iPS cell clones in 96-well plates (scale bar, 100 μ m). **c**, Efficiency of reprogramming obtained from two independent experiments.



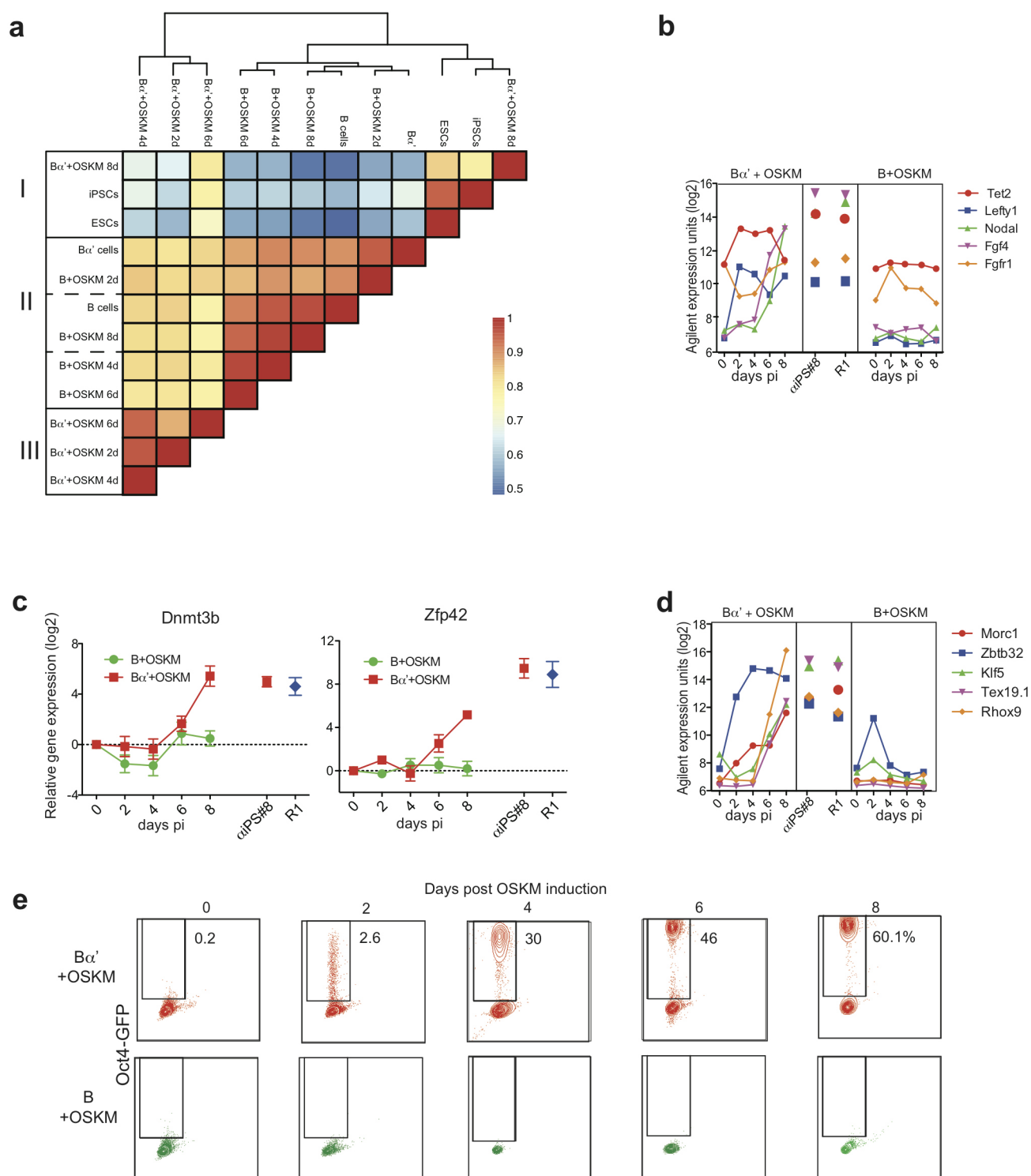
Extended Data Figure 2 | C/EBP α mutant and cell viability during iPS cell reprogramming. **a**, E2 has no effect on reprogramming efficiency and ectopic Oct4 expression. Error bars indicate s.d. ($n = 3$). **b**, Effect of a C/EBP α mutant for the DNA binding domain on iPS cell reprogramming efficiency. Student's t -test *** $P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$).

c, Histogram showing percentages of live cells during the first 3 days of iPS cell reprogramming as analysed by FACS. Dead cells were scored as positive for DAPI staining. **d**, Effect of 18 h C/EBP α pulses after OSKM induction. Error bars indicate s.d. ($n = 3$).



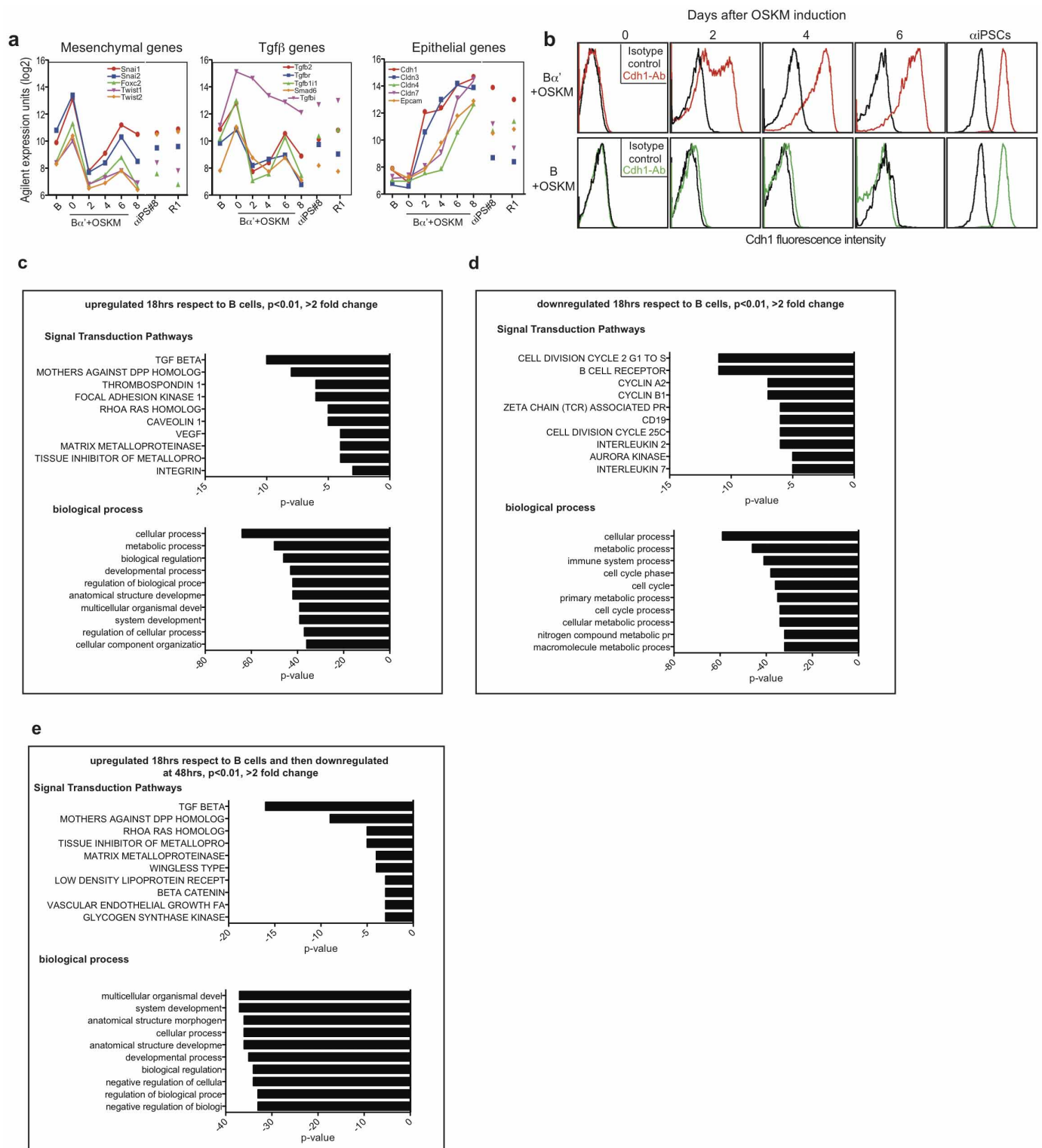
Extended Data Figure 3 | Properties of αiPS cells. **a**, Heat maps showing gene expression profiles of B cells, ES cells, αiPS cells (clone 22) and iPS cells (clone 26) derived from Bα' and B cells, respectively. Genes were selected based on a >4-fold difference between B cells and ES cells. **b**, Micrographs (×10) of αiPS cell clone 22 growing on MEF feeders, showing colonies of ES-like morphology (bright field) and expression of Nanog (green) and Oct4 (red). **c**, Embryoid bodies generated from αiPS cell clone 22 (bright field)

differentiated into Tuj1-positive cells (ectoderm), Foxa2-positive cells (endoderm) and SMA expressing cells (mesoderm). Original magnification, ×10. **d**, Section of a teratoma obtained from αiPS clone 22 stained with haematoxylin and eosin, showing differentiation into the major germ layers. Original magnification, ×10. **e**, Proportion of chimaeric mice obtained after injection of CD1 blastocysts with three different αiPS cell clones.



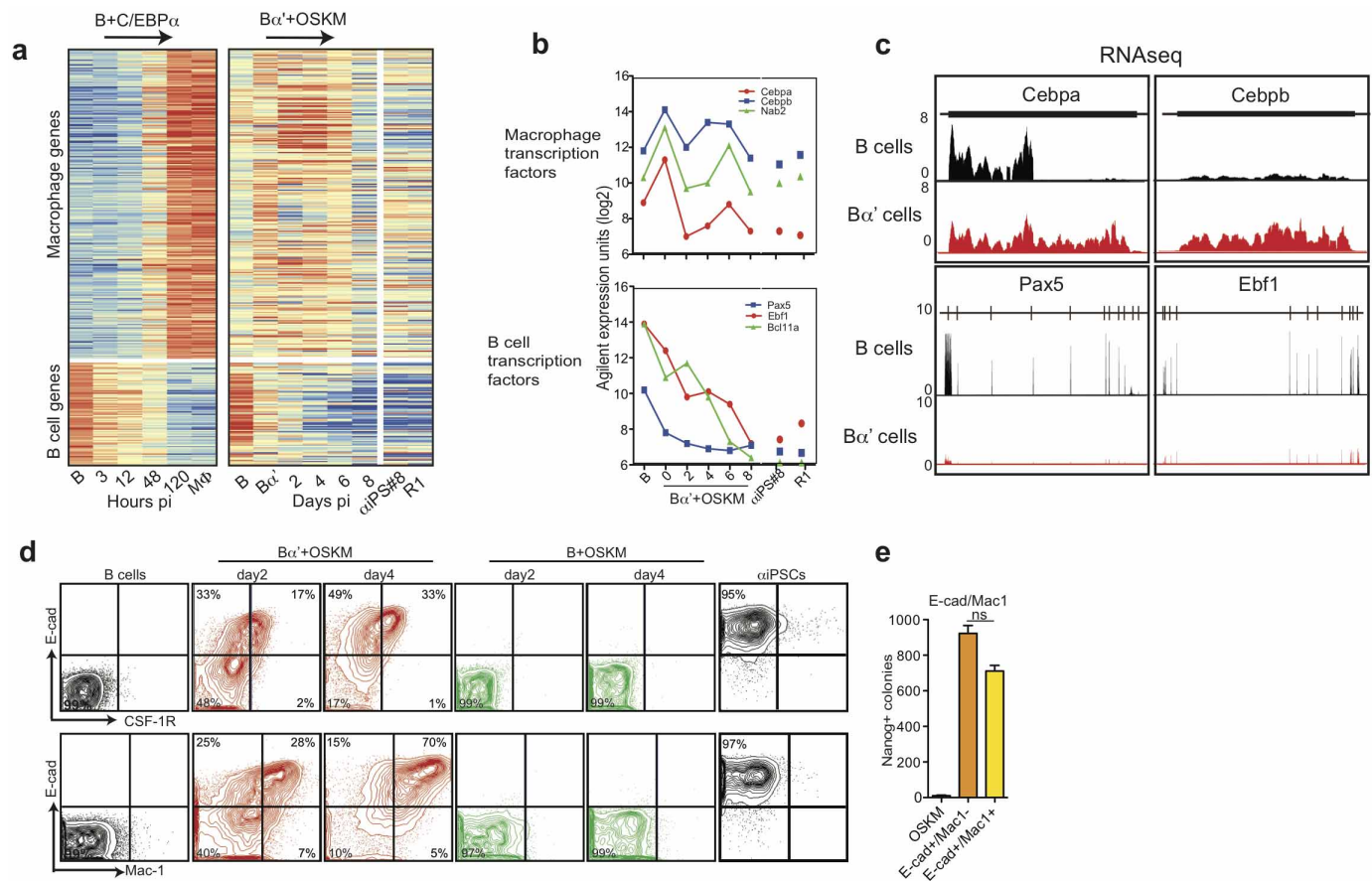
Extended Data Figure 4 | Pluripotency gene regulation during iPS cell reprogramming. **a**, Unsupervised hierarchical clustering analysis of gene expression array data. **b**, Agilent gene expression kinetics (\log_2 expression units) of pluripotency genes from $B\alpha'$ + OSKM cells (red shaded panels), α iPS cells and ES cells (white panels) and B + OSKM cells (green shaded panels).

c, qRT-PCR confirming array results for *Dnmt3b* and *Zfp42* (*Rex1*) genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, Agilent gene expression kinetics of late transition genes. **e**, Oct4-GFP expression during iPS cell reprogramming.



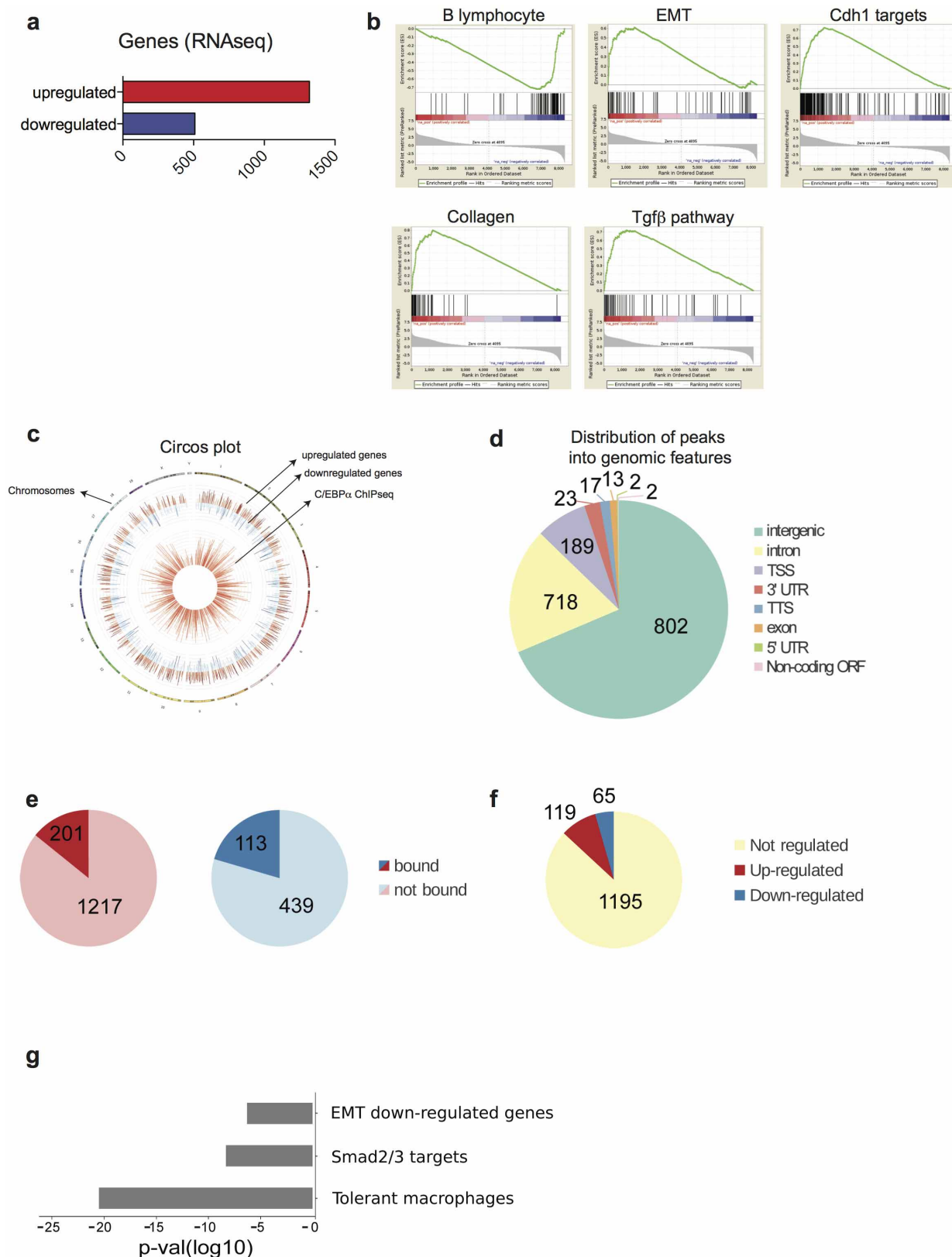
Extended Data Figure 5 | Changes in the expression of mesenchymal-epithelial transition genes. **a**, Agilent gene expression kinetics of mesenchymal-epithelial transition (MET) genes. **b**, E-cadherin expression by FACS at different times after induction. Isotype controls are shown in black; antibody-stained cells in red and green. **c**, GO analysis for genes upregulated

> 2 -fold in $B\alpha'$ cells compared with B cells. **d**, GO analysis for genes downregulated > 2 -fold in $B\alpha'$ cells compared with B cells. **e**, GO analysis of genes upregulated > 2 -fold in $B\alpha'$ cells compared with B cells and subsequently downregulated > 2 -fold at 48 h.



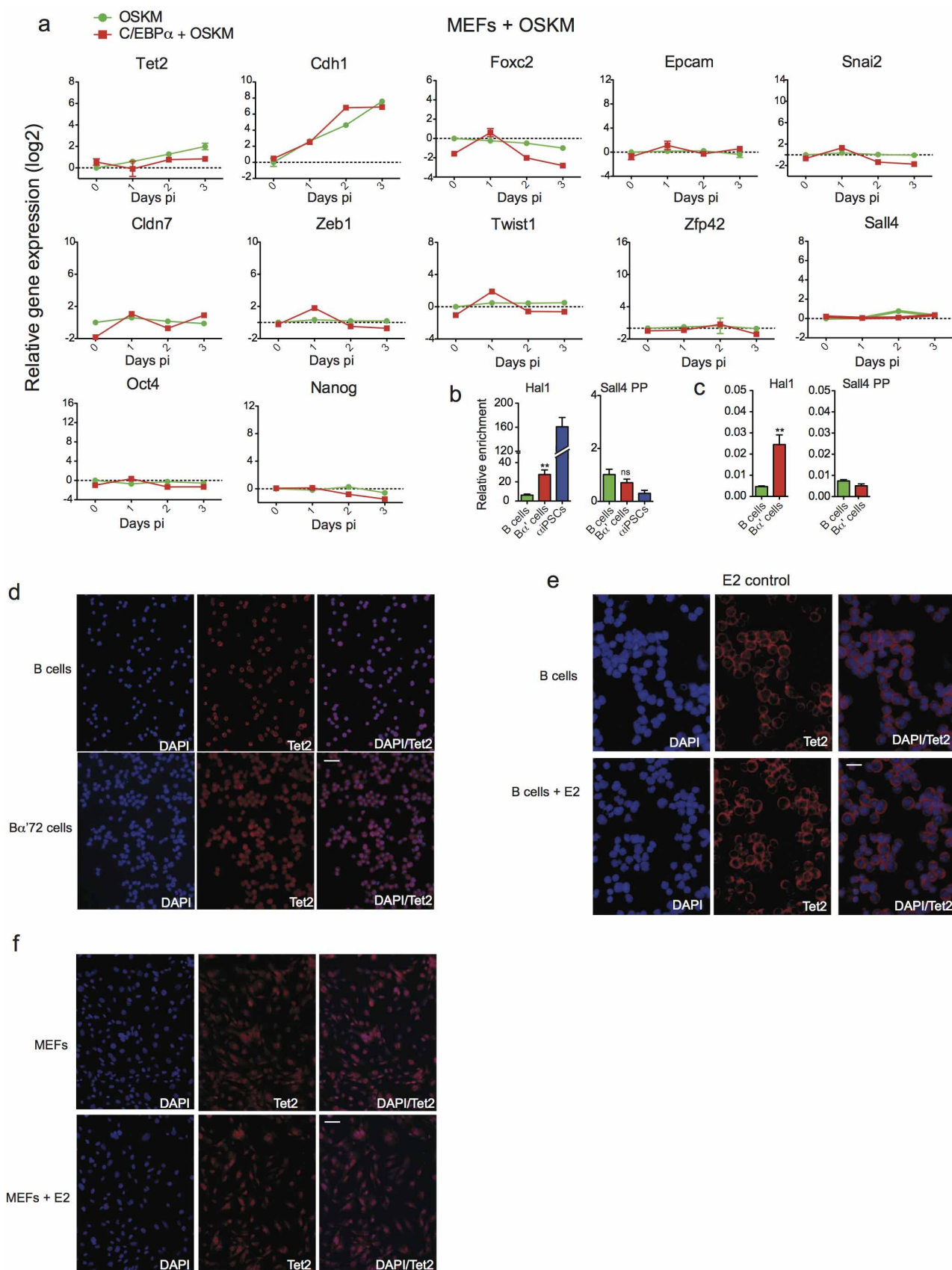
Extended Data Figure 6 | Changes in the expression of B cell and macrophage genes. **a**, Heat map of expression Affymetrix arrays with 263 macrophage and 83 B-cell-specific genes during C/EBP α -induced transdifferentiation of B cells and heat maps of the same genes using data from the Agilent expression array in Fig. 2. **b**, Agilent gene expression kinetics of

selected macrophage and B-cell transcription factors. **c**, RNA-seq data for two representative macrophage and B-cell-specific genes each. **d**, FACS profiles of E-cadherin (Cdh1) expression, combined with either CSF-1R (upper panels) or Mac1 (lower panels) staining. **e**, Reprogramming potential of E-cadherin⁺/Mac1⁻ and E-cadherin⁺/Mac1⁺ cells. Error bars indicate s.d. ($n = 3$).



Extended Data Figure 7 | ChIP-seq and RNA-seq analysis. **a**, Number of upregulated and downregulated genes (>1.5 -fold change) after a pulse of C/EBP α . **b**, Result of gene set enrichment analysis on RNA-seq data (ranked by fold change) for selected significantly enriched gene sets (from MSigDB cp.v4 database). **c**, Circular visualization of genomic distribution of differentially expressed genes (RNA-seq) and C/EBP α binding sites (ChIP-seq). **d**, Distribution of C/EBP α peaks into genomic features, on the basis of RefSeq

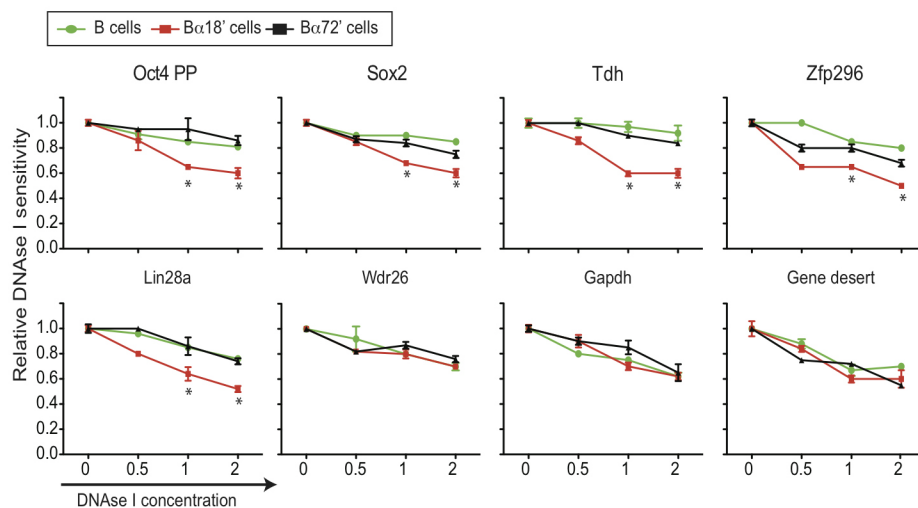
mm9 annotation. **e**, Fraction of upregulated (red chart) and downregulated genes (blue) after a pulse of C/EBP α showing a peak at 50 kb from TSS. **f**, Variation of expression after C/EBP α pulse of nearest genes to peaks. **g**, Selected gene sets significantly enriched in genes adjacent to C/EBP α peaks based on hypergeometric tests (from MSigDB cp.v4 database). Bonferroni procedure, P value <0.05 .



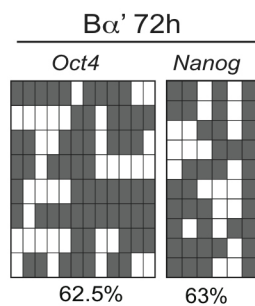
Extended Data Figure 8 | Tet2 expression in B cells and MEFs. **a**, RT-qPCR of Tet2, pluripotency and MET genes after induction of OSKM in MEFs pulsed or not with C/EBP α . Error bars indicate s.d. ($n = 3$). **b**, **c**, hMeDIP and Tet2 binding at regulatory regions of the Tet2 target gene *Hal1* (ref. 3) and *Sall4*.

d, Tet2 localization in B cells pulsed with C/EBP α for 72 h. **e**, Oestradiol treatment has no effect on cellular localization of Tet2 protein in B cells (scale bar, 30 μ m). **f**, Cellular localization of Tet2 protein in MEFs treated or not with E2 (scale bar, 100 μ m).

a



b

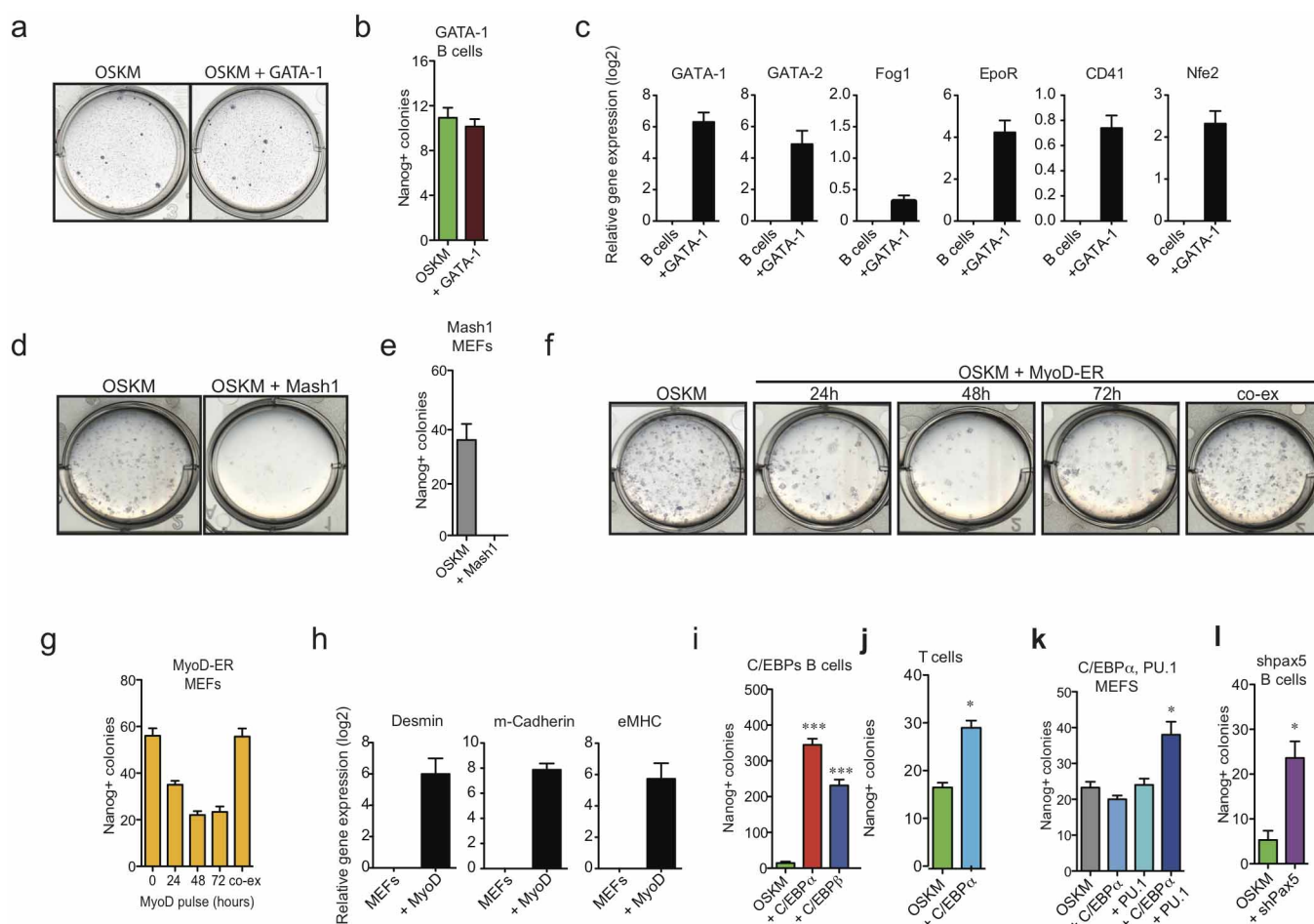


c

	B	Bα'	Bα' 2d	Bα' 4d	Bα' 6d	Bα' 8d	αiPS	Genes
1.wave PP genes								<i>Oct4, Lin28, Tdh, Gdf3, Zfp296</i>
2.wave PP genes								<i>Nanog, Sox2, Esrrb, Sall4, Zfp42, Dppa5a</i>
'Very late' PP genes								<i>Morc1, Zbtb32, Klf5, Tex19.1, Rhox9</i>
Epithelial genes								<i>Cdh1, Cdn3, Cdn4, Cdn7, Epcam</i>
Mesenchymal genes								<i>Snai1, 2, Foxc2, Col1a2, Col7a1, Twist1, 2</i>
TGFβ pathway								<i>Tgfb2, Tgfb1, Tgfb3, Tgfb4, Smad6</i>
MΦ assoc. genes								<i>Cebpa, Cebpb, Nab2</i>
B cell assoc. genes								<i>Pax5, Ebf1, Tcf3, Bcl11a</i>

Extended Data Figure 9 | DNase I sensitivity and gene expression summary. a, DNase I sensitivity at regulatory regions in B cells and Bα' cells (data from Fig. 4a) and B cells pulsed for 72 h determined by enzyme titration. Error bars indicate s.d. ($n = 3$). Student's t -test $*P < 0.05$ relative to control. b, CpGs

analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters. Filled rectangles represent methylated CpGs, empty rectangles unmethylated residues. c, Summary of relevant gene expression changes during the transition from B cells to Bα' cells and Bα' cells to αiPS cells.



Extended Data Figure 10 | Transcription factor and cell type specificity.

a, b, B+OSKM cells were infected with a retrovirus expressing GATA1 or left uninfected, induced with doxycycline and scored for 12 d.p.i. alkaline phosphatase⁺ and Nanog⁺ colonies. Error bars indicate s.d. ($n = 3$). **c**, qRT-PCR for megakaryocyte/erythroid-restricted genes in B cells infected with GATA1 retrovirus for 3 days. Error bars indicate s.d. ($n = 3$). **d, e**, Alkaline phosphatase⁺ colonies at 15 d.p.i. in OSKM MEFs infected with a retrovirus expressing Mash1 or mock infected, and Nanog⁺ colonies counted. Error bars indicate s.d. ($n = 3$). **f, g**, Effect of MyoD expression on iPS cell reprogramming of MEFs. Cells were infected with MyoD-ER and sequentially induced with E2 to activate MyoD and doxycycline to activate OSKM. Representative plates

with alkaline phosphatase⁺ iPS cell colonies 15 d.p.i. of MEFs pre-induced for 24, 48 or 72 h or continuously with E2. Error bars indicate s.d. ($n = 3$).

h, Expression of muscle-restricted genes in OSKM MEFs 3 days after MyoD-ER induction. Error bars indicate s.d. ($n = 3$). **i**, Effects of C/EBPα and C/EBPβ in B+OSKM cells. Student's t -test *** $P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$). **j**, Effect of an 18-h C/EBPα pulse in pre-T+OSKM cells on the formation of 12 d.p.i. Nanog⁺ colonies. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **k**, Effects of C/EBPα, PU.1 and their combination in MEFs. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **l**, Effect of Pax5 knock down in B-cell reprogramming. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$).

C/EBP α poises B cells for rapid reprogramming into induced pluripotent stem cells

Bruno Di Stefano^{1,2}, Jose Luis Sardina^{1,2*}, Chris van Oevelen^{1,2*}, Samuel Collombet^{3,4,5}, Eric M. Kallin^{1,2†}, Guillermo P. Vicent^{1,2}, Jun Lu⁶, Denis Thieffry^{3,4,5}, Miguel Beato^{1,2} & Thomas Graf^{1,2,7}

CCAAT/enhancer binding protein- α (C/EBP α) induces transdifferentiation of B cells into macrophages at high efficiencies and enhances reprogramming into induced pluripotent stem (iPS) cells when co-expressed with the transcription factors Oct4 (Pou5f1), Sox2, Klf4 and Myc (hereafter called OSKM)^{1,2}. However, how C/EBP α accomplishes these effects is unclear. Here we find that in mouse primary B cells transient C/EBP α expression followed by OSKM activation induces a 100-fold increase in iPS cell reprogramming efficiency, involving 95% of the population. During this conversion, pluripotency and epithelial–mesenchymal transition genes become markedly upregulated, and 60% of the cells express Oct4 within 2 days. C/EBP α acts as a ‘path-breaker’ as it transiently makes the chromatin of pluripotency genes more accessible to DNase I. C/EBP α also induces the expression of the dioxygenase Tet2 and promotes its translocation to the nucleus where it binds to regulatory regions of pluripotency genes that become demethylated after OSKM induction. In line with these findings, overexpression of Tet2 enhances OSKM-induced B-cell reprogramming. Because the enzyme is also required for efficient C/EBP α -induced immune cell conversion³, our data indicate that Tet2 provides a mechanistic link between iPS cell reprogramming and B-cell transdifferentiation. The rapid iPS reprogramming approach described here should help to fully elucidate the process and has potential clinical applications.

Reprogramming of somatic cells into induced pluripotent stem (iPS) cells by the OSKM transcription factors (also called Yamanaka factors) can be divided into a stochastic and a deterministic phase^{4,5}. During this process cells undergo a mesenchymal–epithelial transition (MET) and activate endogenous pluripotency genes, paralleled by changes in histone marks, nucleosome positioning and chromatin accessibility⁶. In addition, their promoters become de-methylated, preceded by oxidation of methylated CpGs through the dioxygenases Tet2 and Tet1, both of which have been implicated in the establishment of pluripotency^{7,8}. Earlier work showed that co-expression of C/EBP α with OSKM increases the reprogramming efficiency of B cells ~15-fold, reaching ~3% of the population¹. Here we describe that a pulse of C/EBP α followed by OSKM overexpression permits the rapid reprogramming of B cells into iPS cells by activating Tet2 and facilitating accessibility of pluripotency gene promoters to Oct4 binding. Highly efficient reprogramming of somatic cells has recently also been reported with a loss-of-function approach⁹.

Committed B-cell precursors (hereafter referred to as B cells) can be induced to transdifferentiate into macrophages at 100% efficiency by forced C/EBP α expression, deregulating ~7,500 genes¹⁰. Reasoning that chromatin of cells in transition might be more ‘open’ than that of end stages, we tested the effect of transiently exposing B cells to C/EBP α , followed by OSKM expression. B cells were isolated from the bone marrow of reprogrammable mice (containing a tetracycline-controlled transactivator (rtTA) and a doxycycline-responsive OSKM cassette¹¹),

infected them with C/EBP α -ER-hCD4 retrovirus, sorted human CD4⁺ cells 4 days later and incubated them for different times with β -estradiol (E2) followed by a wash-out (Fig. 1a). Subsequently, OSKM was induced by doxycycline treatment and Nanog⁺ colonies scored 12 days post induction (d.p.i.). B cells continuously co-expressing C/EBP α with OSKM showed an 11-fold enhancement in reprogramming efficiency compared to cells induced with OSKM alone (B+OSKM cells), confirming earlier reports^{1,12}. In contrast, cells treated for 18 h with E2 and then treated with doxycycline (B α ' + OSKM cells) exhibited a 103-fold colony increase, with cells pulsed for 6 h already showing a 74-fold increase (Fig. 1b–d). Clonal assays showed that 92–94% of viable colonies were Nanog⁺ after 12 days (Extended Data Fig. 1a–c). OSKM induction of B cells pre-treated with E2 did not increase Nanog⁺ colony numbers nor Oct4 expression levels (Extended Data Fig. 2a). A mutant of C/EBP α (BRM-2 (ref. 13)) defective for DNA binding failed to enhance iPS cell generation (Extended Data Fig. 2b). Induced B α ' + OSKM cells remained >50% viable (Extended Data Fig. 2c). C/EBP α pulses after OSKM induction had no effect on reprogramming efficiency (Extended Data Fig. 2d), indicating that C/EBP α acts as a path-breaker for OSKM-induced reprogramming. Stable iPS cell lines derived from B α ' + OSKM cells (α iPS cells) displayed similar gene expression profiles as embryonic stem (ES) cells, differentiated into all three germ layers *in vitro* and *in vivo*, and efficiently contributed to coat colour chimaerism (Fig. 1e and Extended Data Fig. 3a–e).

B α ' + OSKM-derived iPS cell colonies could be identified as early as 4 d.p.i. and their numbers increased modestly after 8 days (Fig. 1f, g), whereas B+OSKM colonies continued to increase 8–10 d.p.i. (Fig. 1f, g). Retroviral hCD4 expression was found to be silenced within 2–4 days in B α ' + OSKM cells compared to ~8 days in B+OSKM cells (Fig. 1h). To test the effect of C/EBP α on OSKM-induced transgene independence, B α ' and B cells were doxycycline-treated for different times (Fig. 1i). The first transgene-independent iPS cell colonies were observed after 4 days for B α ' + OSKM cells, compared to 9–10 days for B+OSKM controls (Fig. 1j). Together these results show that the C/EBP α pulse accelerates iPS cell reprogramming by 4–6 days.

Gene expression analyses of B α ' + OSKM cells 8 d.p.i. showed the upregulation of 764 out of 1,668 genes expressed more highly in ES cells than in B cells, including all well described pluripotency genes (Fig. 2a). Unsupervised hierarchical clustering analysis revealed that 8 d.p.i. B α ' + OSKM cells clustered with ES/iPS cells (Extended Data Fig. 4a). A large part of known pluripotency genes were activated 2–6 d.p.i. (Fig. 2b, c and Extended Data Fig. 4b), reaching levels comparable to α iPS cells and ES cells within ~1 week (Fig. 2c and Extended Data Fig. 4c). In contrast, they remained essentially silent in B+OSKM cells (Fig. 2b, c). In addition, five genes described to be activated very late (21 d.p.i.) during fibroblast reprogramming¹⁴ became upregulated within 2–4 days in B α ' + OSKM cells (Extended Data Fig. 4d). Using Oct4–GFP reporter mice¹² crossed with reprogrammable mice¹¹, B α ' + OSKM

¹Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003 Barcelona, Spain. ²Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain. ³Ecole Normale Supérieure, Institut de Biologie de l'ENS, 45 Rue d'Ulm, Paris F-75005, France. ⁴Inserm, U1024, Paris F-75005, France. ⁵CNRS, UMR 8197, Paris F-75005, France. ⁶Yale Cancer Center, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg Lluís Companys 23, 08010 Barcelona, Spain. [†]Present address: Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York 10065, USA.

*These authors contributed equally to this work.

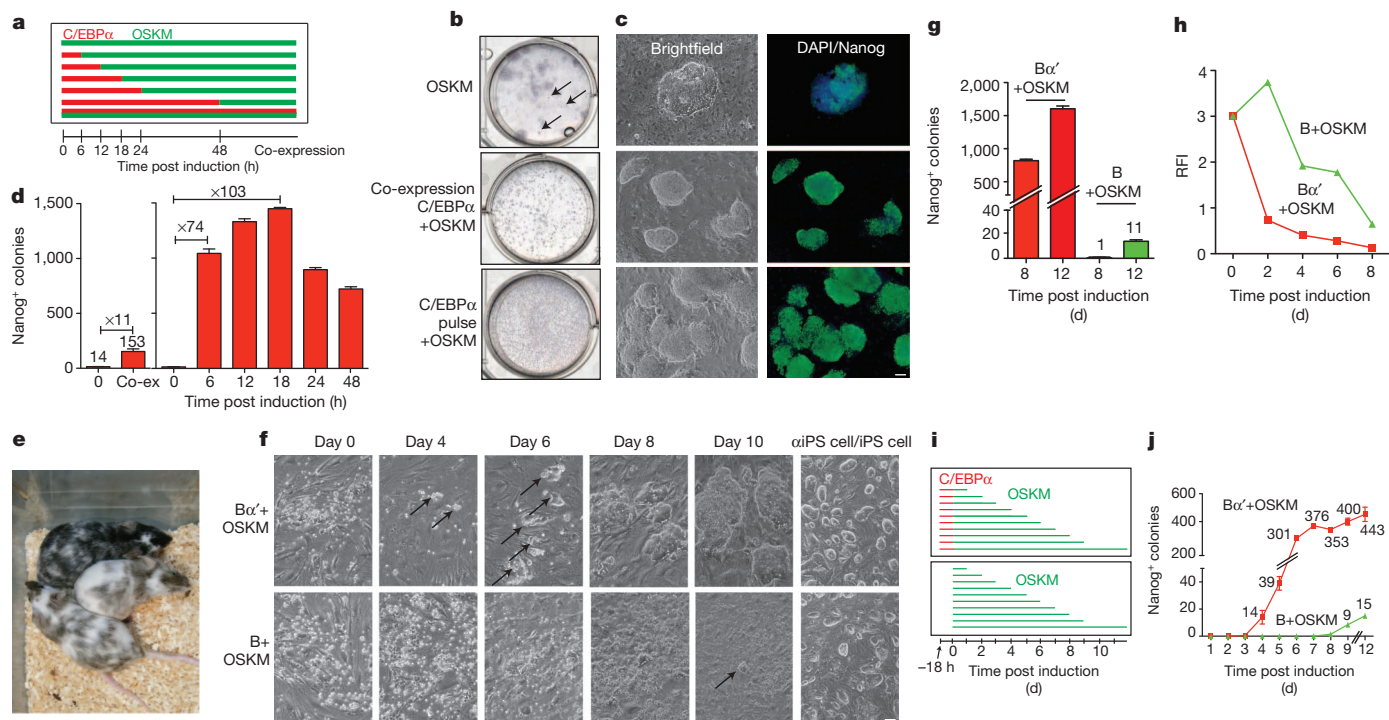


Figure 1 | Effect of transient C/EBP α expression on OSKM-induced iPS cell reprogramming of B cells. **a**, Experimental strategy. **b**, Six-well plates with alkaline-phosphatase-positive iPS cell colonies at 12 d.p.i. obtained from B-cell cultures treated with doxycycline to induce OSKM only (B + OSKM), cells co-expressing C/EBP α with OSKM, or cells treated for 18 h with E2 to activate C/EBP α followed by OSKM induction (B α' + OSKM). **c**, Nanog $^{+}$ iPS cell colonies at 12 d.p.i. (scale bar, 100 μ m). **d**, Quantification of 12 d.p.i. Nanog $^{+}$ colonies after OSKM induction in B cells pre-treated with C/EBP α for the indicated pulse lengths. Error bars indicate s.d. ($n = 3$). **e**, Chimeric mice

obtained after blastocyst injection of α iPS cell clone 12. **f**, Doxycycline-induced B α' + OSKM cultures with arrows indicating earliest detectable ES-cell-like colonies (scale bar, 100 μ m). **g**, Nanog $^{+}$ colonies of B α' + OSKM and B + OSKM cultures at 8 and 12 d.p.i. Error bars indicate s.d. ($n = 3$). **h**, Retroviral silencing after induction of B α' + OSKM and B + OSKM cells measured by hCD4 expression. RFI, relative fluorescence intensity. **i**, **j**, Experimental strategy to determine OSKM transgene independence (**i**) and formation of Nanog $^{+}$ iPS cell colonies (**j**) at 12 d.p.i. Error bars indicate s.d. ($n = 3$).

cells became 30% GFP $^{+}$ at 4 d.p.i. and 60% GFP $^{+}$ at 8 d.p.i., whereas B + OSKM cells remained negative (Extended Data Fig. 4e). Serum-free conditions further accelerated the process with 62% of the B α' + OSKM cells becoming GFP $^{+}$ already at 2 d.p.i. and 95% at 4 d.p.i. (Fig. 2d, e), with a similar viability as in serum containing medium. Induced B + OSKM cells died without serum.

We next scored the expression of MET genes^{15,16}. B α' + OSKM cells upregulated all epithelial genes tested and expressed E-cadherin 2–4 d.p.i., again with B + OSKM cells remaining negative (Extended Data Fig. 5a, b). In line with the possibility that the C/EBP α pulse is capable of initiating an epithelial–mesenchymal transition (EMT)¹⁷, mesenchymal genes encoding TGF- β pathway members, transcription factors and collagens were first upregulated by the C/EBP α pulse and then switched off after OSKM induction (Extended Data Fig. 5a, e). To compare gene expression changes during OSKM-induced reprogramming of B α' cells with C/EBP α -induced transdifferentiation of B cells we selected genes selectively expressed in macrophages (263) and B cells (83) (Supplementary Table 1). Most of these genes were already up- or down-regulated in B α' cells, respectively, as expected from the induced conversion (Extended Data Fig. 6a). After OSKM induction, B-cell genes became further silenced whereas macrophage genes elicited a more heterogeneous response (Extended Data Fig. 6b, c). B α' + OSKM cells at 2–4 d.p.i. co-expressed E-cadherin and the macrophage markers CSF-1R and Mac1, whereas B + OSKM cells remained negative and α iPS cells expressed exclusively E-cadherin (Extended Data Fig. 6d). Sorted Mac1 $^{+}$ E-cadherin $^{+}$ cells and Mac1 $^{-}$ E-cadherin $^{+}$ cells yielded similar proportions of iPS cell colonies, indicating that expression of Mac1 has no role in iPS cell reprogramming (Extended Data Fig. 6e). Analysis of the 18 h pulsed B cells by RNA-seq revealed 1,418 upregulated and 552 downregulated genes (Extended Data Fig. 7a). Gene set enrichment analysis (GSEA)

showed a strong correlation between genes downregulated in B α' cells with sets of B-cell-specific genes, and upregulated genes with genes related to EMT, collagen expression and the TGF- β pathway (Extended Data Fig. 7b), confirming the array data. C/EBP α chromatin immunoprecipitation sequencing (ChIP-seq) analysis revealed 1,766 DNA regions bound by the protein in B α' cells of which \sim 20% were within 50 kb of genes whose expression changed (Extended Data Fig. 7c–g). These included two regulatory sites in *Tet2* (Fig. 3a, ref. 3) as well as sites close to lymphoid, MET and TGF- β pathway genes. *Tet2* became upregulated \sim 1.6-fold after the 18-h pulse and was further activated by OSKM, reaching levels comparable to ES cells within 2 days (Fig. 3b, c). In contrast, mouse embryonic fibroblasts (MEFs) pulsed with C/EBP α did not upregulate the gene (Extended Data Fig. 8a). Tet2 overexpression enhanced iPS cell reprogramming of B cells \sim 3-fold (Fig. 3d), supporting the reported role of Tet2 in iPS cell formation^{7,8}. Because the Tet2-induced 5-hydroxymethylcytosine (5hmC) modification has been described to participate in the maintenance of pluripotency, we determined 5hmC levels by hydroxymethylated DNA immunoprecipitation followed by qPCR (hMeDIP-qPCR) in B α' cells at regulatory regions of nine pluripotency genes for which hydroxymethylation levels change during reprogramming¹⁸. 5hmC levels were significantly increased for all genes tested (Fig. 3e and Extended Data Fig. 8b) and Tet2 binding was detected at the same regions (Fig. 4f and Extended Data Fig. 8c). Unexpectedly, Tet2 was predominantly cytoplasmic in uninduced B cells and nuclear after the C/EBP α pulse (Fig. 3g). The nuclear localization was maintained after sustained expression of C/EBP α (72 h) (Extended Data Fig. 8d). No protein translocation was observed in B cells treated with E2 alone or infected with a control vector (Extended Data Fig. 8e) and MEFs exhibited predominantly nuclear Tet2 localization (Extended Data Fig. 8f). Because hydroxymethylation is required

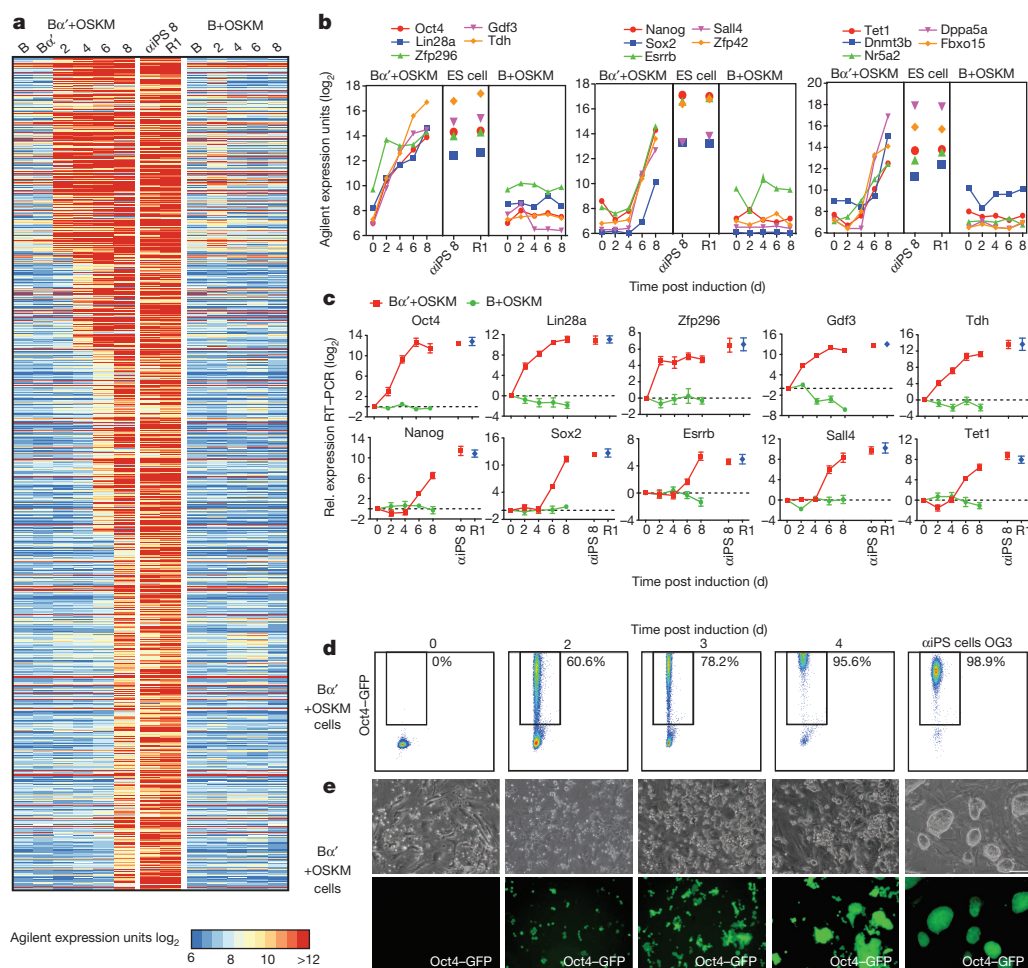


Figure 2 | Upregulation of pluripotency genes. **a**, Heat map of Agilent expression arrays showing upregulation of genes expressed >4-fold higher in ES cells than in B cells at the indicated time points after doxycycline treatment of B α' +OSKM cells and B+OSKM cells, with α iPS cells (clone 8) and ES cells (R1) as controls. **b**, Array expression kinetics of three groups of pluripotency

genes from B α' +OSKM cells, α iPS cells and ES cells and B+OSKM cells. **c**, qRT-PCR validation of selected genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, **e**, Oct4-GFP expression analysed by FACS and microscopy in B α' +OSKM cells (scale bar, 100 μ m).

for active DNA demethylation, we analysed the DNA methylation status of the *Oct4* and *Nanog* promoters using bisulphite sequencing (Fig. 3h). Whereas the *Oct4* promoter became partially demethylated in B α' +OSKM as early as 2 d.p.i. (Fig. 3h), the *Nanog* promoter only became demethylated at 8 d.p.i., consistent with the observed delayed upregulation of the gene (Fig. 2b, c).

The 5hmC changes at pluripotency genes raised the possibility that the C/EBP α pulse eliminates epigenetic barriers, increasing chromatin accessibility to the OSKM transcription factors. To test this, we performed quantitative DNase I sensitivity assays on selected regulatory regions of pluripotency genes. This revealed increased accessibility to the enzyme of all sites tested in the 18-h pulsed B cells whereas control regions and pluripotency sites in control B+OSKM cells showed no difference (Fig. 4a). Interestingly, after prolonged induction of C/EBP α (72 h) the chromatin of pluripotency genes returned to the closed state as seen in untreated B cells (Extended Data Fig. 9a). In addition, no demethylation of the *Nanog* and *Oct4* promoters was observed (Extended Data Fig. 9b). We next tested Oct4 occupancy in OSKM-induced B α' cells at sites within enhancers and promoters of pluripotency genes bound by Oct4 in ES cells¹⁹ (Fig. 4b). At 2 d.p.i. all of these sites showed significant Oct4 occupancy, whereas they remained unbound in B+OSKM cells (Fig. 4b). Together, these findings show that the C/EBP α pulse transiently induces chromatin remodelling at regulatory regions of pluripotency genes, making them rapidly accessible to Oct4 binding.

Our results show that C/EBP α poises B cells for highly efficient and almost immediate reprogramming into iPS cells by OSKM transcription factors, with MET and pluripotency network activation occurring concomitantly (summarized in Extended Data Fig. 9c). In addition, our data indicate that the C/EBP α pulse already initiates an EMT, reminiscent of the recently described EMT-MET transition during reprogramming of fibroblasts¹⁷. The single cell cloning and Oct4-GFP reporter experiments suggest that essentially 100% of the poised cells can be reprogrammed into iPS cells. Several lines of evidence suggest that the rapid demethylation of the *Oct4* promoter involves Tet2: the C/EBP α pulse directly binds to and upregulates the *Tet2* gene and shuttles the protein into the nucleus. Then Tet2 binds to regulatory sites of *Oct4* and other pluripotency genes, converting 5mC residues into 5hmC (summarized in Fig. 4c). This modification might directly cause the gene's de-repression or be linked to demethylation and/or induction of chromatin remodelling⁸. The effect of C/EBP α is highly specific as only the closely related factor C/EBP β , but not the lineage-instructive factors GATA1 (ref. 20), nor Mash1 or MyoD expressed in fibroblasts, was active (Extended Data Fig. 10a–i). C/EBP α also sensitized T-cell precursors²¹ but not fibroblasts (Extended Data Fig. 10j, k). However, when C/EBP α was co-expressed with PU.1, a condition that mediates transdifferentiation to macrophage-like cells²², it induced a ~6-fold enhancement of iPS cell reprogramming, whereas PU.1 alone had no effect (Extended Data Fig. 10k). This indicates that a myeloid determinant sensitizes somatic cells for iPS cell reprogramming, in line with

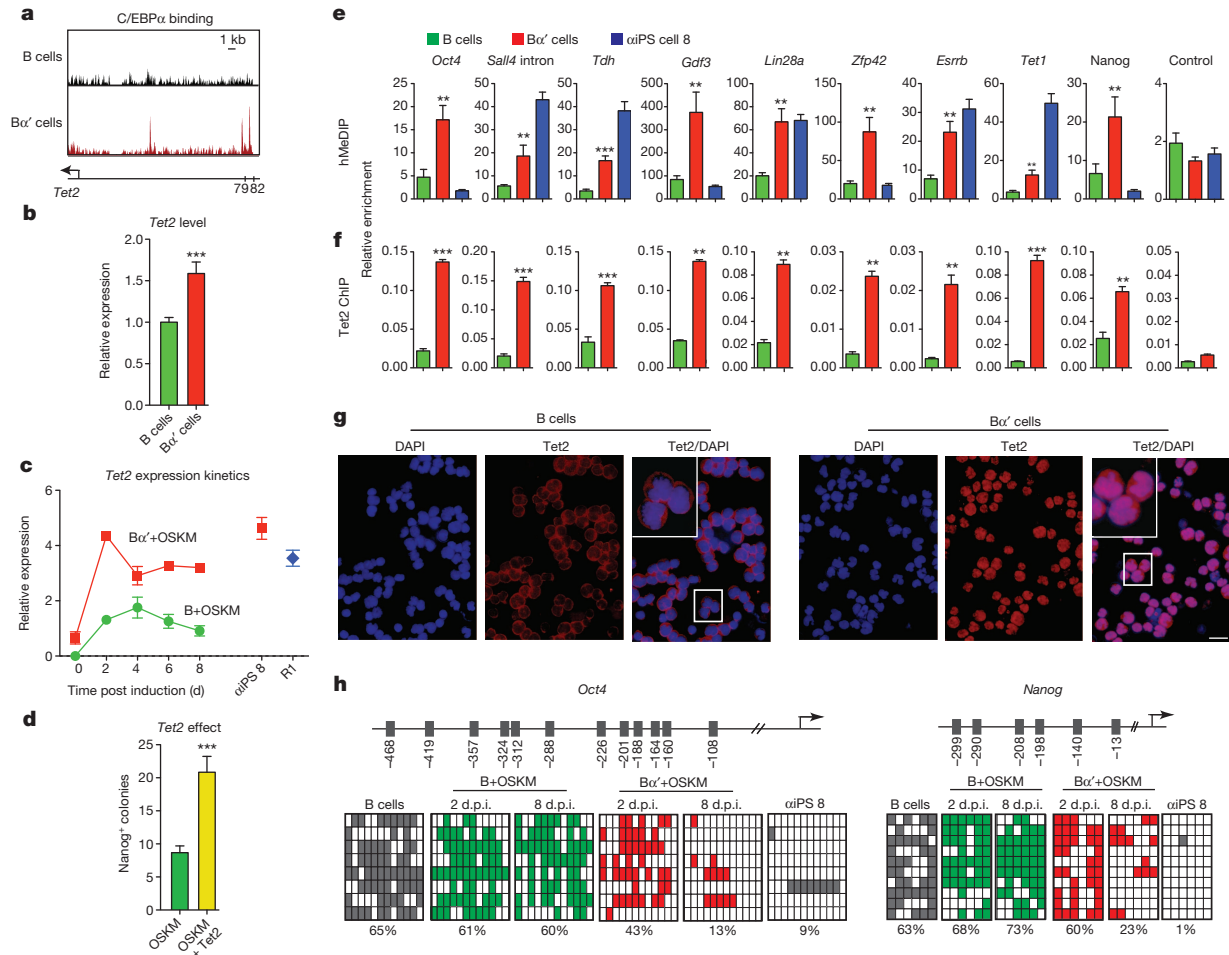


Figure 3 | Tet2 activation, cytosine hydroxymethylation and DNA methylation. **a**, C/EBPα binding at two enhancers upstream of the *Tet2* gene. **b**, c, qRT-PCR for *Tet2* expression. Error bars indicate s.d. (n = 6). Student's *t*-test ****P* < 0.001 relative to control. **d**, *Tet2* overexpression effect on B-cell reprogramming. Error bars indicate s.d. (n = 6). Student's *t*-test ****P* < 0.001 relative to control. **e**, hMeDIP followed by qPCR at regulatory regions of selected pluripotency genes. Error bars indicate s.d. (n = 3). Student's *t*-test

P* < 0.01, *P* < 0.001 relative to B cells. **f**, Tet2 binding sites determined by ChIP of the regions in **e**. Error bars indicate s.d. (n = 3). Student's *t*-test ***P* < 0.01, ****P* < 0.001 relative to B cells. **g**, Immunofluorescence of *Tet2* protein in B cells and Bα' cells (scale bar, 30 μm). **h**, CpGs analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters, with their position indicated on the top. Filled rectangles represent methylated CpGs; empty rectangles represent unmethylated residues.

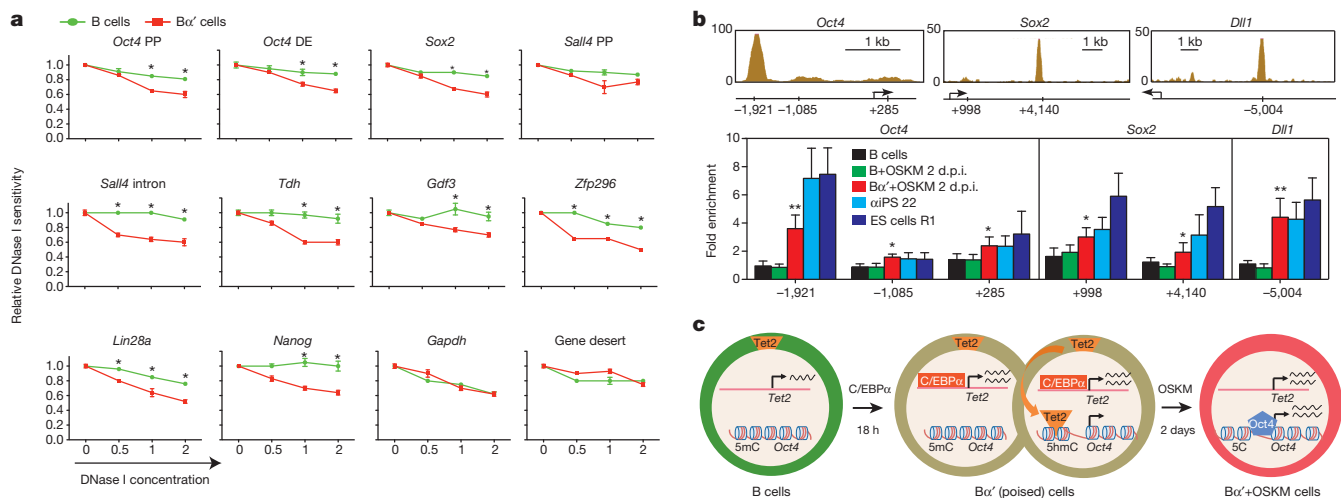


Figure 4 | Chromatin accessibility and summary schemes. **a**, DNase I sensitivity at regulatory regions in B cells and Bα' cells determined by enzyme titration. Error bars indicate s.d. (n = 3). Student's *t*-test **P* < 0.05 relative to control. DE, distal enhancer; PP, proximal promoter. **b**, Oct4 binding sites examined at the *Pou5f1*, *Sox2* and *Dll1* loci, with Oct4 peaks corresponding to ES cells¹⁹. Oct4 binding by ChIP-qPCR analysis. Student's *t*-test **P* < 0.05, ***P* < 0.01 relative to control. Error bars indicate s.d. (n = 3). **c**, Model of reprogramming mechanism involving *Tet2*.

examined at the *Pou5f1*, *Sox2* and *Dll1* loci, with Oct4 peaks corresponding to ES cells¹⁹. Oct4 binding by ChIP-qPCR analysis. Student's *t*-test **P* < 0.05, ***P* < 0.01 relative to control. Error bars indicate s.d. (n = 3). **c**, Model of reprogramming mechanism involving *Tet2*.

the finding that macrophage/granulocyte progenitors, the formation of which is driven by C/EBP α ²³, show the highest described reprogramming efficiency of somatic cells (25%)¹². The C/EBP α effect might recapitulate an embryonic function of the gene as it is required, in combination with C/EBP β , for the formation of functional trophoblasts²⁴, cells with intriguing similarities to macrophages²⁵.

METHODS SUMMARY

Somatic cells used for iPS cell reprogramming were derived from the reprogrammable mouse¹¹. ES cells and iPS cells were cultured on mitomycin-C-treated MEFs in ES cell medium containing 15% FBS and 1,000 U ml⁻¹ of LIF. For B- and T-cell reprogramming experiments the medium was supplemented with cell-type-specific cytokines¹². OSKM expression was induced with 2 ng ml⁻¹ of doxycycline. α iPS cells and iPS cells correspond to individual iPS cell clones obtained after expansion into stable lines of OSKM induced B α' and B cells, respectively. The total number of iPS cell colonies was scored after alkaline phosphatase or Nanog staining. For qRT-PCR and microarray analyses, B cells at different time points during reprogramming, as well as ES cells and iPS cells, were trypsinized and FACS sorted to remove feeder cells and dead cells. Total RNA was extracted using the miRNeasy kit (Qiagen). Microarrays were done using Agilent 8X60K expression arrays. Details on injections into blastocysts to test for chimaerism are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 April; accepted 15 November 2013.

Published online 15 December 2013.

- Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Kallin, E. M. *et al.* Tet2 facilitates the derepression of myeloid target genes during CEBP α -induced transdifferentiation of pre-B cells. *Mol. Cell* **48**, 266–276 (2012).
- Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
- Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
- Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835–850 (2011).
- Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374 (2013).
- Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
- Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
- Di Tullio, A. *et al.* CCAAT/enhancer binding protein α (C/EBP α)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108**, 17016–17021 (2011).
- Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods* **7**, 56–59 (2010).
- Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41**, 968–976 (2009).
- Porse, B. T. *et al.* E2F repression by C/EBP α is required for adipogenesis and granulopoiesis *in vivo*. *Cell* **107**, 247–258 (2001).
- Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012).
- Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
- Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
- Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nature Cell Biol.* **15**, 829–838 (2013).
- Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
- Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
- Kulesa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblats, and erythroblasts. *Genes Dev.* **9**, 1250–1262 (1995).
- Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP α and PU.1 transcription factors. *Immunity* **25**, 731–744 (2006).
- Feng, R. *et al.* PU.1 and C/EBP α / β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
- Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α -deficient mice. *Proc. Natl Acad. Sci. USA* **94**, 569–574 (1997).
- Begay, V., Smink, J. & Leutz, A. Essential requirement of CCAAT/enhancer binding proteins in embryogenesis. *Mol. Cell. Biol.* **24**, 9744–9751 (2004).
- Guilbert, L., Robertson, S. A. & Wegmann, T. G. The trophoblast as an integral component of a macrophage-cytokine network. *Immunol. Cell Biol.* **71**, 49–57 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Batlle for generating the chimaeric mice, M. Stadtfeld, D. Egli, P. Cosma, G. Filion, B. Lehner and J. Valcarcel for critical reading of the manuscript, V. Broccoli, P. Muñoz and S. Aznar Benitah for constructs and antibodies and H. Schoeler for the Oct4 reporter mice. This work was supported by the Ministerio de Educacion y Ciencia, SAF.2007-63058 and AGAUR 2009 SGR768. B.D.S. holds a La Caixa International PhD Fellowship.

Author Contributions T.G. and B.D.S. conceived the idea for this project, designed the experiments and wrote the paper. B.D.S. performed the cell culture, animal experiments and molecular biology and B.D.S. and J.L.S. the methylation analyses. J.L.S. performed the studies on Tet2 and J.L.S. and C.v.O. the ChIP experiments. S.C., E.M.K. and D.T. were responsible for the bioinformatics and G.P.V. and M.B. for the DNase I experiments and analyses. J.L. provided the Tet2 overexpressing plasmids.

Author Information The array data are available from the Gene Expression Omnibus under the accession number GSE52397. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.G. (Thomas.Graf@crg.eu).

METHODS

Mice. The reprogrammable mouse containing a tetracyclin-inducible OSKM cassette as well as rTA has been described previously¹¹.

Cell cultures. ES cells and iPS cells were cultured on mitomycin-C-treated MEF feeder cells in KO-DMEM (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, non-essential amino acids, β -mercaptoethanol, 1,000 U ml⁻¹ LIF (ES cell medium) and 15% fetal bovine serum (FBS, Invitrogen).

MEF cultures were established by trypsin digestion of mouse embryos (embryonic day 13.5) and the resulting cells cultured in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin. CD19⁺ pro-B and pre-B cells were isolated from bone marrow using monoclonal antibodies to CD19 (1D3), obtained from BD Pharmingen, using MACS (Miltenyi Biotech). Double-negative (DN) T cells were prepared from thymii of 4-week-old mice by lineage depletion with Streptavidin microbeads. The following biotin-conjugated antibodies were used to label lineage (lin)-positive thymocytes before lineage depletion: CD3 (145-2C11), CD4 (GK1.5), CD8 (53-6.7), B220 (RA3-6B2), CD19 (1D3), Mac1/CD11b (M1/70), Gr1 (RB6-8C5), Ter119 (TER-119), I-A/I-E (2G9). The purity of the sorted cell fractions was confirmed by FACS using an LSR II machine (BD). After isolation, B cells were expanded in RPMI medium supplemented with 10% FBS and IL-7 for 1 week before inducing reprogramming. T cells were grown RPMI medium supplemented with 10% FBS supplemented with IL-7, Flt3 and SCF.

Reprogramming. Reprogramming experiments were conducted in gelatinized plates seeded with a feeder layer of the OP9 stromal cell line, using ES cell medium supplemented with 2 μ g ml⁻¹ of doxycycline and 15% FBS. For the reprogramming of B cells, IL-4 (10 ng ml⁻¹), IL-7 (10 ng ml⁻¹) and IL-15 (2 ng ml⁻¹) were added to the medium (ref. 13). B cells were seeded at a density of 500 cells cm⁻² in 6-well plates. For the conditioning of B α' cells, C/EBP α -infected sorted B cells were exposed for 18 h to 100 nM of β -estradiol (E2) followed by inducer wash-out and addition of ES cell medium supplemented with doxycycline and cytokines. Serum-free iPS reprogramming was performed in serum-free ES cell medium (KO-DMEM supplemented with 15% knockout serum replacement, L-glutamine, penicillin/streptomycin, non-essential amino acids, β -ME, 1,000 U ml⁻¹ LIF, N2 (100X), B27 (100X))⁹ supplemented with 2 μ g ml⁻¹ doxycycline to activate OSKM factors. 21 inhibitors were added to the medium 2 days after doxycycline induction.

For expansion of iPS cell lines, colonies with ES cell morphology were picked after doxycycline withdrawal at 12 d.p.i. OSKM MEFs were seeded on mitomycin-C-treated MEF feeders in ES cell medium containing FBS and induced with doxycycline.

Vectors and virus production and infection. C/EBP α -ER-GFP, C/EBP α -ER-CD4, C/EBP β , PU.1, C/EBP α -BRM-2 retroviruses and shPax5 lentivirus have been described previously^{1,2,13}. *Gata1* cDNA was cloned by PCR in the BglII/XhoI restriction sites of the MIG vector. The MyoD-ER plasmid was obtained from P. Muñoz, Mash1 from V. Broccoli. The retroviral vector for murine Tet2 overexpression was derived from MSCV-based pMIRWAY-puro-TET2²⁶ by replacing puro with EGFP. Virus production has been described previously²⁷. Briefly, HEK293T cells were co-transfected with vector plasmid and packaging plasmids using calcium phosphate transfection. Viral supernatants were collected 48–72 h later and concentrated by ultracentrifugation at 20,000g for 2 h at 20 °C. Viral concentrates were re-suspended in 1 \times PBS and stored at –80 °C. Infection of MEFs was carried out in medium containing 5 μ g ml⁻¹ polybrene, and infection of B cells by centrifugation with concentrated virus for 2 h at 32 °C at 1000g in B-cell medium.

Alkaline phosphatase staining. For alkaline phosphatase staining, the cells were fixed in 2% paraformaldehyde and then incubated for 15 min in NTMT solution (NaCl 100 mM, Tris A 1 M, Tris B 1 M, MgCl₂ 50 mM, Tween 0.1%) supplemented with BCIP e NBT (Roche).

Immunofluorescence assays. The cells were fixed with 4% paraformaldehyde, blocked and incubated with primary antibodies overnight at 4 °C. On the next day, the cells were exposed to secondary antibodies (all Alexa Fluor from Invitrogen) at room temperature for 1 h. The primary antibodies used were Nanog (Calbiochem), Oct4 (SC-5279), SMA (obtained from P. Muñoz), Tuj1 (obtained from V. Broccoli), Foxa2 (obtained from S. Aznar Benitah) and Tet2 (SC-61). Nuclear staining was performed with DAPI (Invitrogen).

Differentiation of iPS cells. Embryoid bodies (EBs) were derived by plating iPS cells at a concentration of 1.3 \times 10⁶ cells ml⁻¹ in non-adherent dishes in ES medium without LIF. After 4 days in suspension, cell aggregates were plated on gelatin-coated dishes in DMEM medium supplemented with 10% FBS for 10 days.

Teratoma assay. One million iPS cells were injected subcutaneously into SCID BEIGE mice. Three weeks after the injection, the tumours were surgically removed and embedded in paraffin. Teratomas were sectioned into 10- μ m-thick slices using a cryostat (Leica, CM1850 UV) and sections were stained with haematoxylin and eosin.

Chimaeric mice. For the chimaera formation assay, 10–15 iPS cells (C57BL/6J background, black coat colour) were injected into a 3.5 days post coitum (d.p.c.) blastocyst of CD1 mice (white coat colour) and transferred into pseudo-pregnant

2.5 d.p.c. CD-1 recipients for chimaera generation. Chimaerism was ascertained after birth by the appearance of black coat colour (from the C57BL/6J iPS cells) over the white background from the host pups.

FACS. Cell suspensions were stained with various antibodies from either BD Pharmingen or eBiosciences. Cells were analysed with an LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar). Primary antibodies used were CD19, Mac1, B220, Cdh1, CD115 (all from BD Biosciences), hCD4 (eBioscience) and Nanog (Calbiochem).

RNA isolation and quantification. RNA isolation of MEFs and B cells was done with the miRNeasy mini kit (Qiagen). RNA was collected from sorted live cells at 0, 2, 4, 6 and 8 d.p.i. To remove the feeders, ES cells and iPS cell clones at passage 2 or higher were seeded on gelatinized plates and processed with the above kits. RNA was eluted from the columns using RNase-free water or TE buffer and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystem).

Gene expression arrays and qRT-PCR. RNA samples (with an RNA integrity number (RIN) greater than 9) were subjected to transcriptional analyses using Agilent expression arrays. For hybridization, 500 ng of total RNAs were labelled using Agilent's QuickAmp labelling kit following manufacturer instructions and analysed using Agilent 8X60K expression arrays. For the expression data with multiple probes annotated to the same gene only the most dynamic probe as defined by all array hybridizations was considered. Genes with expression changes during reprogramming were defined as showing >4-fold differences between ES cells and uninduced B cells, and having a Student's *t*-test *P* value <0.05 between technical duplicates. B-cell- and macrophage-specific genes were defined by combining array data from a previous study²⁸. Lineage-specific genes were defined as >4-fold upregulated in one lineage over all of the others. Macrophage-specific genes were selected on the basis of >4-fold upregulation in primary macrophages and 5-day C/EBP α -induced macrophages as compared to B cells.

Unsupervised clustering was performed on Pearson correlation coefficients calculated between the indicated array samples using all genes that changed expression at least 2-fold between B cells and ES cells (*n* = 10,982). Data were analysed and visualized using R v2.15.1.

The array data are available from the Gene Expression Omnibus under the accession number GSE52397.

qRT-PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystem) and primers as listed in Supplementary Table 2. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 58 °C and 30 s at 72 °C.

Chromatin immunoprecipitation. ChIP experiments were performed as described previously²⁹. Antibodies against Oct4 (SC-8626) and C/EBP α (SC-61) were purchased from Santa Cruz Biotechnologies. Tet2 ChIP was performed as in ref. 3. Data were obtained from three biological replicates, three independent immunoprecipitations and three technical triplicates.

For ChIP-seq analysis, reads were mapped onto mouse mm9 genome using bowtie2 (v2.1.0) (parameter: –very-sensitive). Reads filtering was done using SAMtools to keep reads that map only once, with a quality score of 10 or more, and to remove duplicates. Peaks were called using HOMER (v4.3) (parameter: style factor) and peaks were selected to control the false discovery rate at 0.001. Peaks annotation and genes association (Extended Data Fig. 7b–d) was performed using HOMER and BedTools (v 2.17).

Circular plot (Extended Data Fig. 7e) was done using Circos.

The sequencing data are available under the Gene Expression Omnibus number GSE52397.

DNA methylation. DNA was extracted using the Blood & Cell Culture DNA Mini kit (Qiagen). Bisulphite treatment of DNA was achieved using the EpiTect bisulphite kit (Qiagen) according to the manufacturer's instructions. The resulting modified DNA was amplified by polymerase chain reaction (PCR) using primer listed in Supplementary Table 2. The resulting amplified products were gel-purified (Qiagen), subcloned into the pGEM-T Vector Systems (Promega), and sequenced using the T7 and SP6 primers.

DNase I assay. Chromatin samples obtained as described before from two biological replicates were subjected to DNase I digestion. Briefly, 2 μ g of chromatin were treated with 0.5, 1, and 2 units of DNase I (Roche) for 3 min at 37 °C in 1 \times DNase incubation buffer. Control samples were incubated in the absence of DNase I. Reactions were terminated by adding EDTA (40 mM final concentration) and the crosslinking was reversed by incubating the samples at 65 °C. After 6 h, proteinase K (40 μ g ml⁻¹ final concentration) was added to each reaction and incubated overnight at 37 °C. After careful phenol-chloroform extractions, the DNA was quantified and used as template for Real Time-PCR reactions using specific primers.

hMeDIP. DNA extraction and IP (DIP) was carried out by a previously published protocol (<http://www.epigenome-noe.net/>, PROT33) with the following alterations.

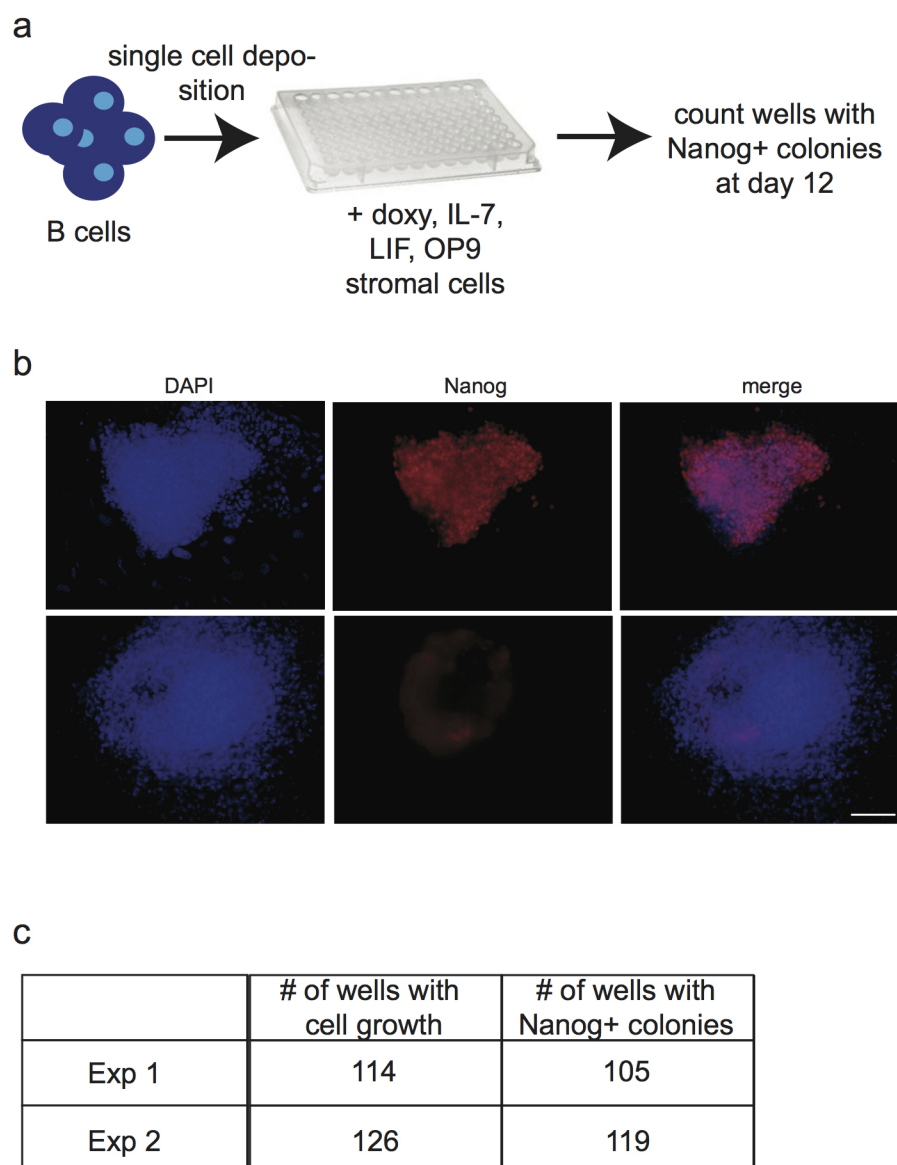
Genomic DNA was fragmented to a mean size of 350 bp using a Bioruptor (Diagenode) for 7×30 s cycles. DNA (re-suspended in 1× IP buffer) was incubated with 5 µg of anti-OHMeC antibody (Active Motif) at 4 °C for 2 h, and 40 µl of a 1:1 mixture of BSA-blocked Protein A (Millipore 16-125) and Protein G (Millipore 16-266) agarose beads were added, followed by an additional 2 h incubation at 4 °C. After washing 3 times in 1× IP buffer, immune complexes were released by incubation with 70 µg Proteinase K, and DNA was extracted once with phenol, extracted once with chloroform, and recovered by EtOH precipitation for qPCR analysis. The 5hmC enrichment in the samples was assessed by qPCR and the values were normalized against *Gapdh* promoter.

RNA-seq. RNA-seq was performed as described previously³⁰. For RNA-seq analysis, reads were aligned onto mouse mm9 genome using STAR and Refseq mm9 annotation for splicing. Reads were filtered to keep only uniquely mapped reads with a maximum of two mismatches. Gene read count was performed using HTseq-count (parameter: mode = union, stranded, features = exons, attribute = gene_id) with Refseq mm9 annotation, and data were normalized using Bioconductor DESeq library (parameter: method = blind, sharingMode = fit-only, fitType = local). Genes with a read count above 300 in all conditions (set as a minimal threshold

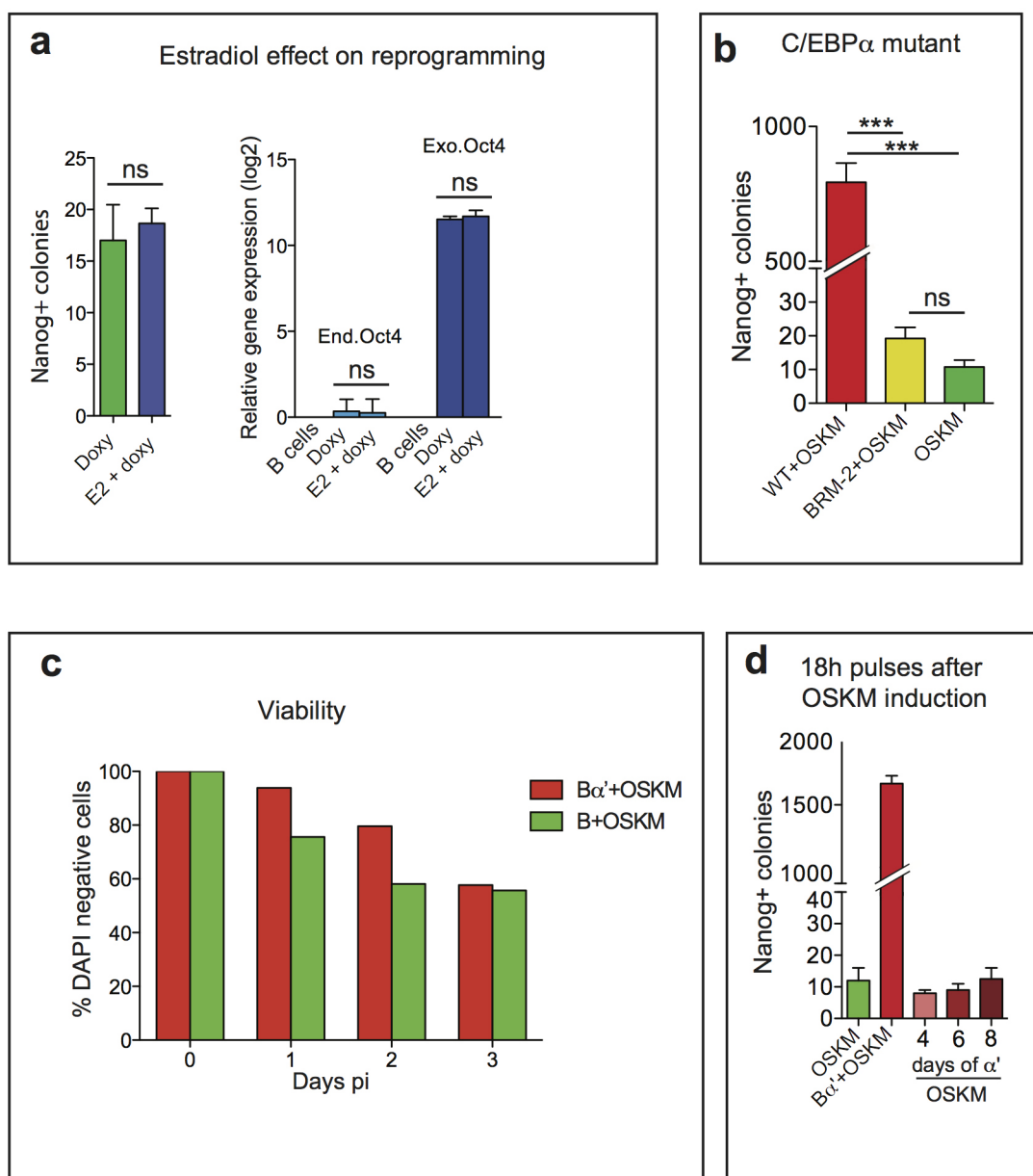
for functional expression) and genes with a 1.5-fold difference in normalized reads count between B cells and B α' cells were considered as differentially expressed. The sequencing data are available under the Gene Expression Omnibus number GSE52397.

Statistical analysis and replicates. All data presented are representative of at least three independent experiments that yielded similar results. Statistical analyses were performed using Prism software (GraphPad).

26. Cheng, J. *et al.* An extensive network of TET2-targeting microRNAs regulates malignant hematopoiesis. *Cell Rep.* **5**, 471–481 (2013).
27. Di Stefano, B. *et al.* A microRNA-based system for selecting and maintaining the pluripotent state in human induced pluripotent stem cells. *Stem Cells* **29**, 1684–1695 (2011).
28. Di Tullio, A. & Graf, T. C/EBP α bypasses cell cycle-dependency during immune cell transdifferentiation. *Cell Cycle* **11**, 2739–2746 (2012).
29. van Oevelen, C. *et al.* A role for mammalian Sin3 in permanent gene silencing. *Mol. Cell* **32**, 359–370 (2008).
30. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).

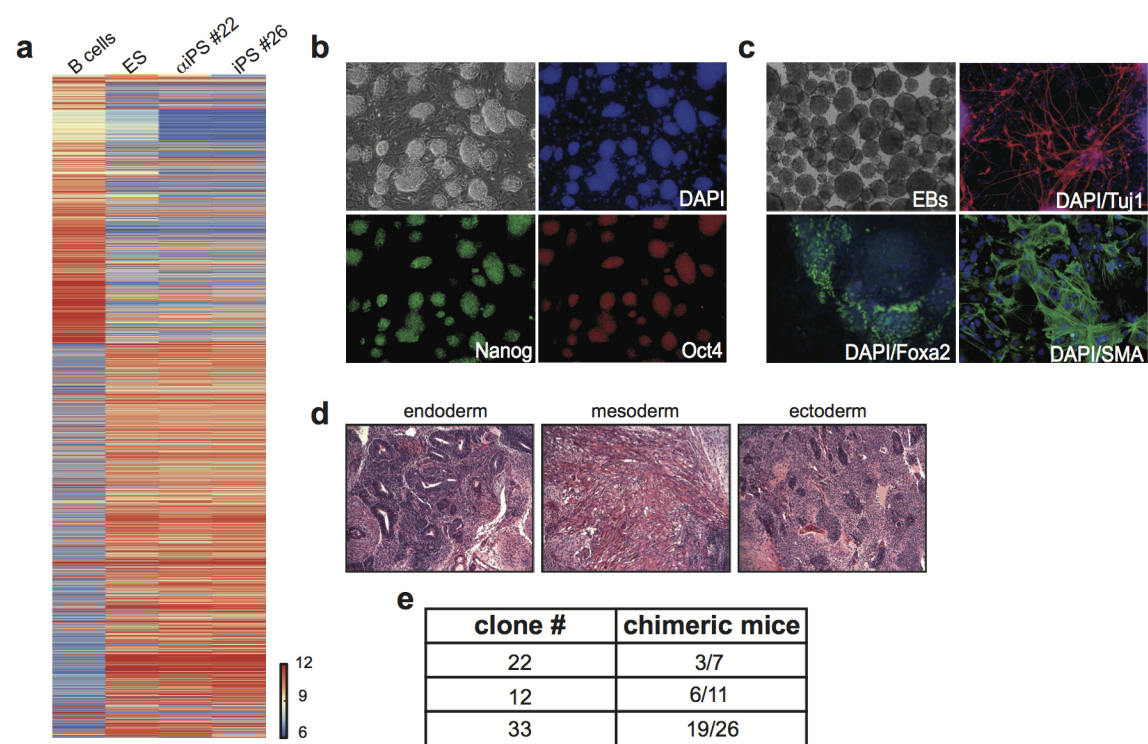


Extended Data Figure 1 | Single cell reprogramming. **a**, B α' cells were plated as single cells in 96-well plates on OP9 feeders in medium conditioned with IL-7, IL-15 and IL-4 cytokines. **b**, Representative Nanog-positive (upper panels) and Nanog-negative (lower panels) iPS cell clones in 96-well plates (scale bar, 100 μ m). **c**, Efficiency of reprogramming obtained from two independent experiments.



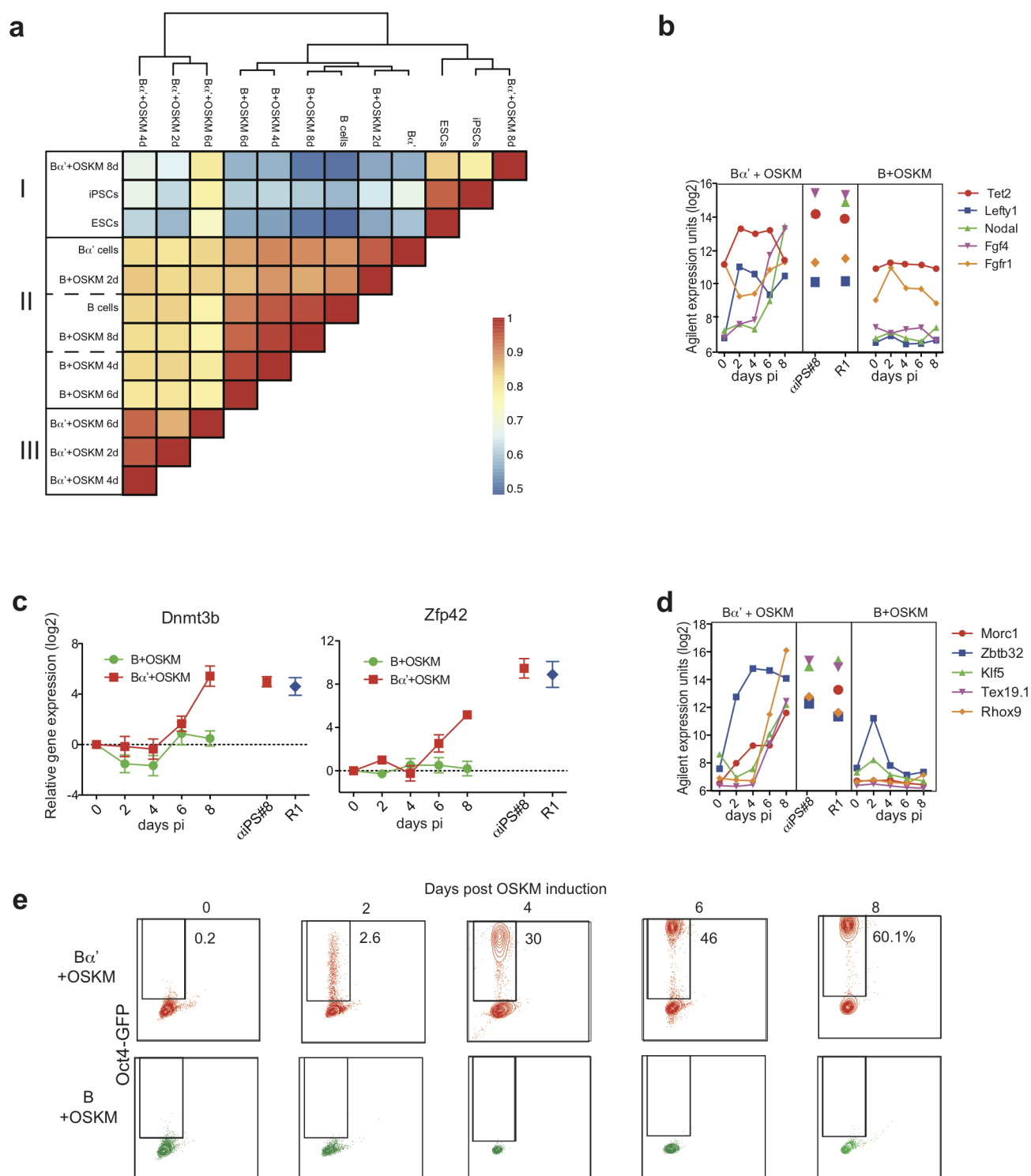
Extended Data Figure 2 | C/EBP α mutant and cell viability during iPS cell reprogramming. **a**, E2 has no effect on reprogramming efficiency and ectopic Oct4 expression. Error bars indicate s.d. ($n = 3$). **b**, Effect of a C/EBP α mutant for the DNA binding domain on iPS cell reprogramming efficiency. Student's t -test *** $P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$).

c, Histogram showing percentages of live cells during the first 3 days of iPS cell reprogramming as analysed by FACS. Dead cells were scored as positive for DAPI staining. **d**, Effect of 18 h C/EBP α pulses after OSKM induction. Error bars indicate s.d. ($n = 3$).



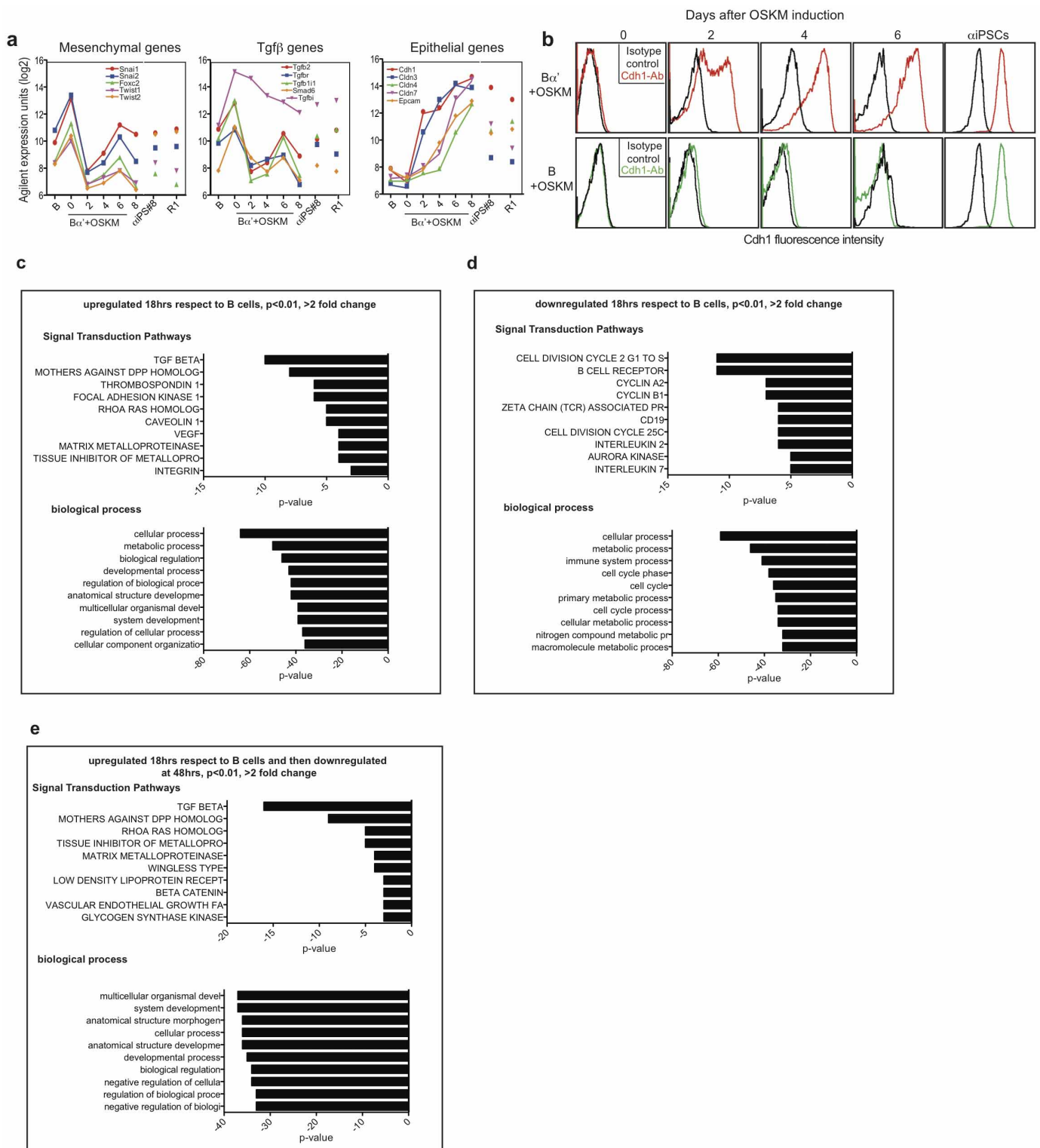
Extended Data Figure 3 | Properties of αiPS cells. **a**, Heat maps showing gene expression profiles of B cells, ES cells, αiPS cells (clone 22) and iPS cells (clone 26) derived from Bα' and B cells, respectively. Genes were selected based on a >4-fold difference between B cells and ES cells. **b**, Micrographs (×10) of αiPS cell clone 22 growing on MEF feeders, showing colonies of ES-like morphology (bright field) and expression of Nanog (green) and Oct4 (red). **c**, Embryoid bodies generated from αiPS cell clone 22 (bright field)

differentiated into Tuj1-positive cells (ectoderm), Foxa2-positive cells (endoderm) and SMA expressing cells (mesoderm). Original magnification, ×10. **d**, Section of a teratoma obtained from αiPS clone 22 stained with haematoxylin and eosin, showing differentiation into the major germ layers. Original magnification, ×10. **e**, Proportion of chimaeric mice obtained after injection of CD1 blastocysts with three different αiPS cell clones.



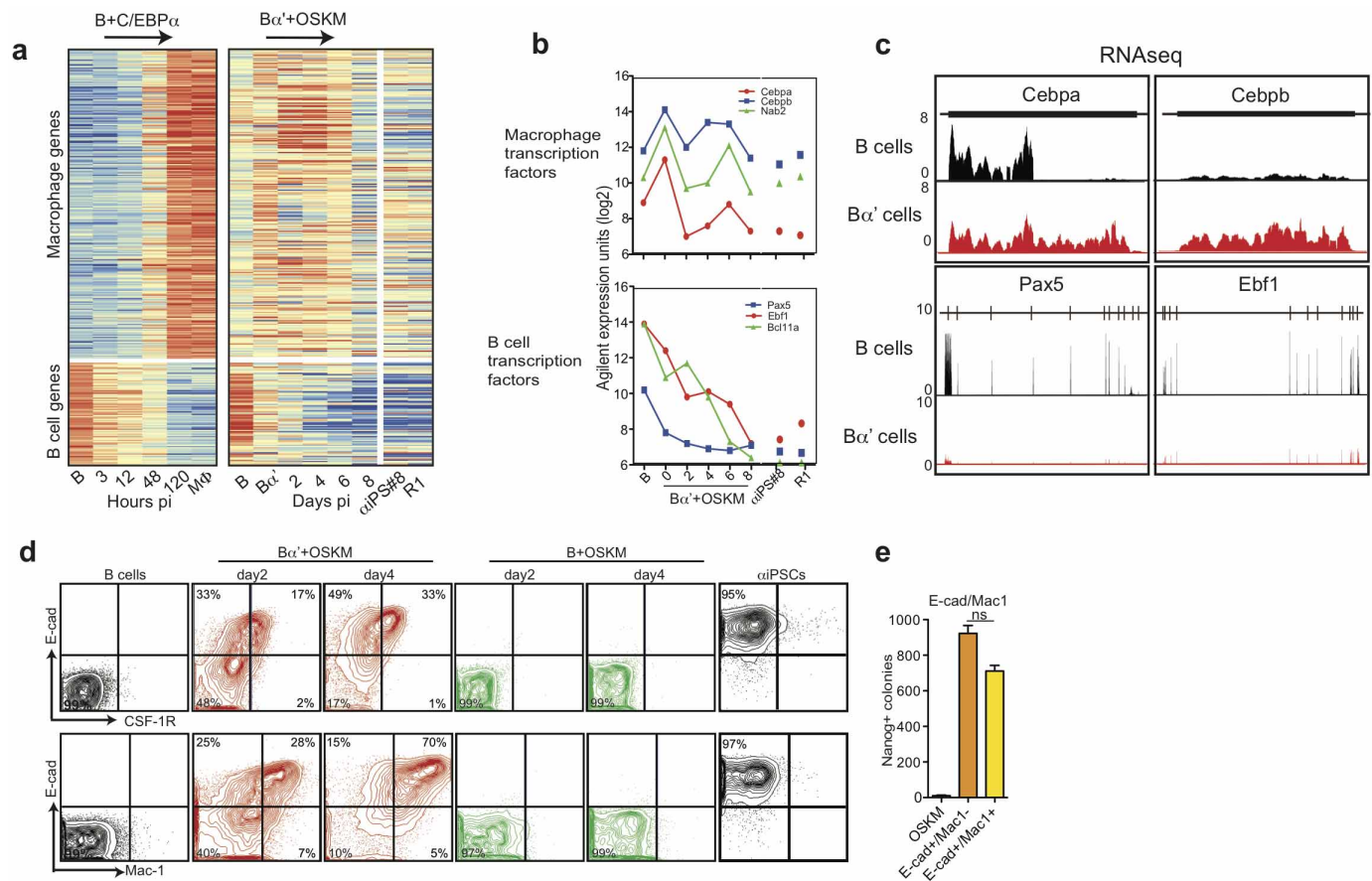
Extended Data Figure 4 | Pluripotency gene regulation during iPS cell reprogramming. **a**, Unsupervised hierarchical clustering analysis of gene expression array data. **b**, Agilent gene expression kinetics (\log_2 expression units) of pluripotency genes from $B\alpha'$ + OSKM cells (red shaded panels), α iPS cells and ES cells (white panels) and B + OSKM cells (green shaded panels).

c, qRT-PCR confirming array results for *Dnmt3b* and *Zfp42* (*Rex1*) genes. Normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$). **d**, Agilent gene expression kinetics of late transition genes. **e**, Oct4-GFP expression during iPS cell reprogramming.



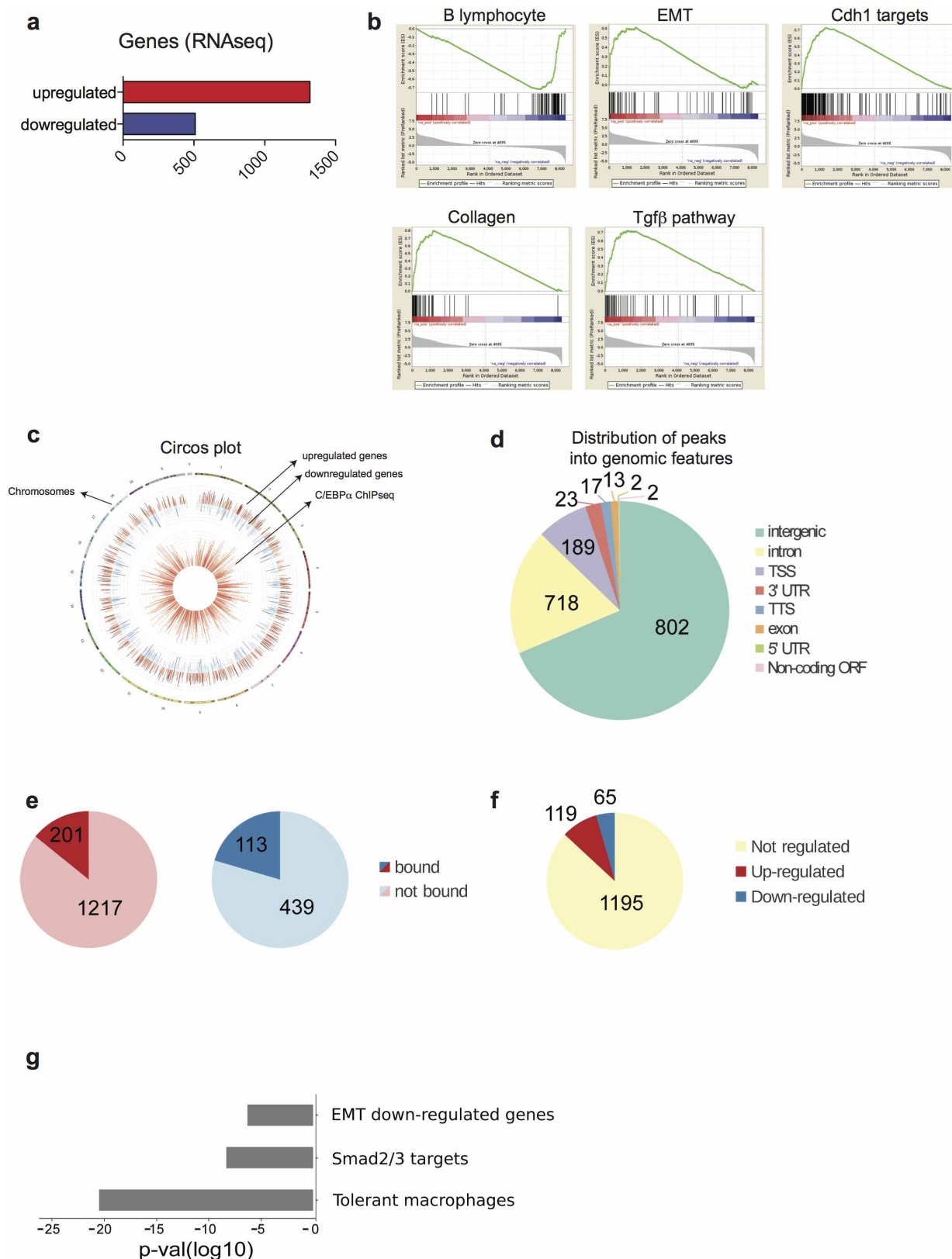
Extended Data Figure 5 | Changes in the expression of mesenchymal-epithelial transition genes. **a**, Agilent gene expression kinetics of mesenchymal-epithelial transition (MET) genes. **b**, E-cadherin expression by FACS at different times after induction. Isotype controls are shown in black; antibody-stained cells in red and green. **c**, GO analysis for genes upregulated

> 2 -fold in $B\alpha'$ cells compared with B cells. **d**, GO analysis for genes downregulated > 2 -fold in $B\alpha'$ cells compared with B cells. **e**, GO analysis of genes upregulated > 2 -fold in $B\alpha'$ cells compared with B cells and subsequently downregulated > 2 -fold at 48 h.



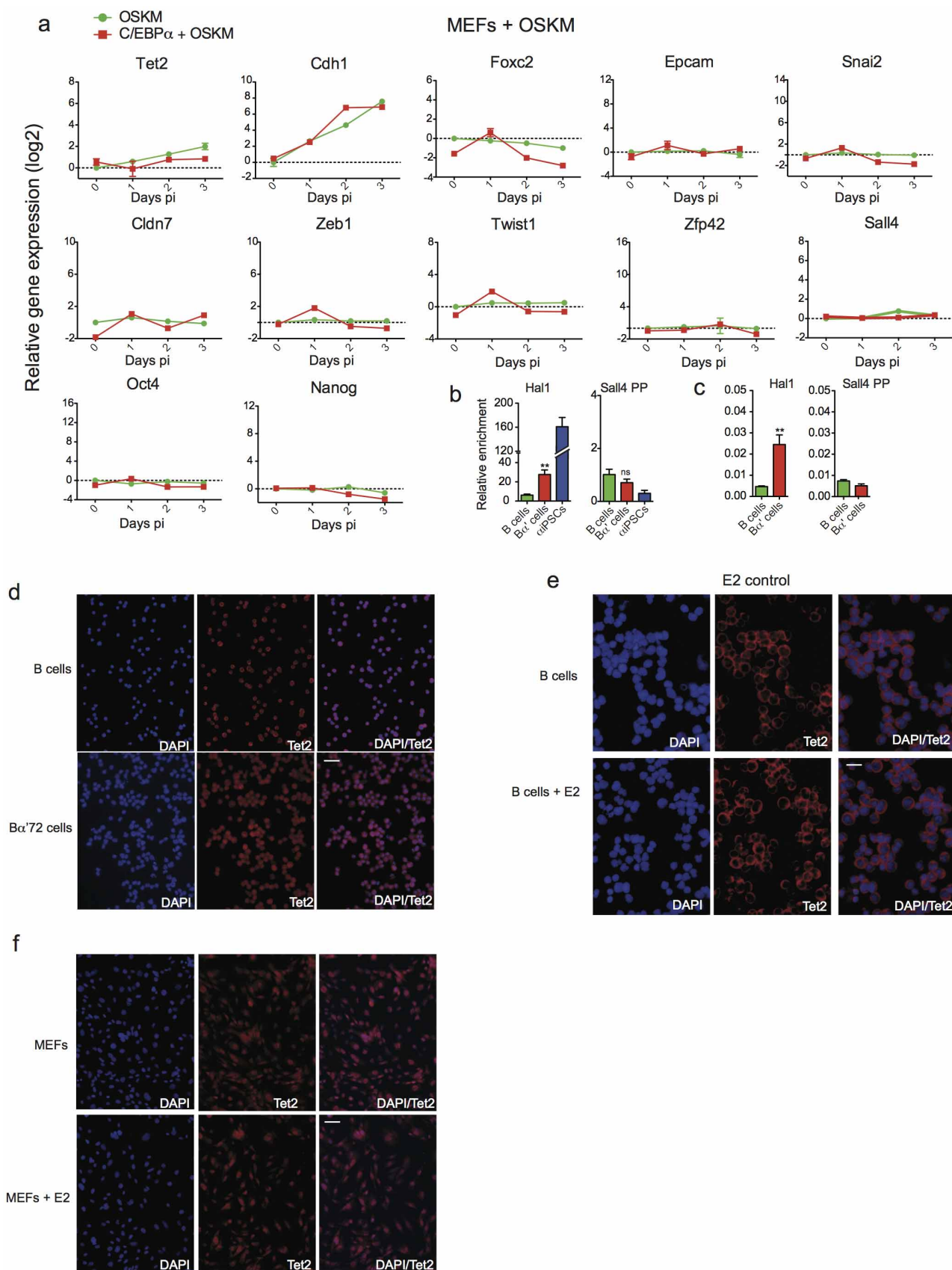
Extended Data Figure 6 | Changes in the expression of B cell and macrophage genes. **a**, Heat map of expression Affymetrix arrays with 263 macrophage and 83 B-cell-specific genes during C/EBP α -induced transdifferentiation of B cells and heat maps of the same genes using data from the Agilent expression array in Fig. 2. **b**, Agilent gene expression kinetics of

selected macrophage and B-cell transcription factors. **c**, RNA-seq data for two representative macrophage and B-cell-specific genes each. **d**, FACS profiles of E-cadherin (Cdh1) expression, combined with either CSF-1R (upper panels) or Mac1 (lower panels) staining. **e**, Reprogramming potential of E-cadherin⁺/Mac1⁻ and E-cadherin⁺/Mac1⁺ cells. Error bars indicate s.d. ($n = 3$).



Extended Data Figure 7 | ChIP-seq and RNA-seq analysis. **a**, Number of upregulated and downregulated genes (>1.5 -fold change) after a pulse of C/EBP α . **b**, Result of gene set enrichment analysis on RNA-seq data (ranked by fold change) for selected significantly enriched gene sets (from MSigDB cp.v4 database). **c**, Circular visualization of genomic distribution of differentially expressed genes (RNA-seq) and C/EBP α binding sites (ChIP-seq). **d**, Distribution of C/EBP α peaks into genomic features, on the basis of Refseq

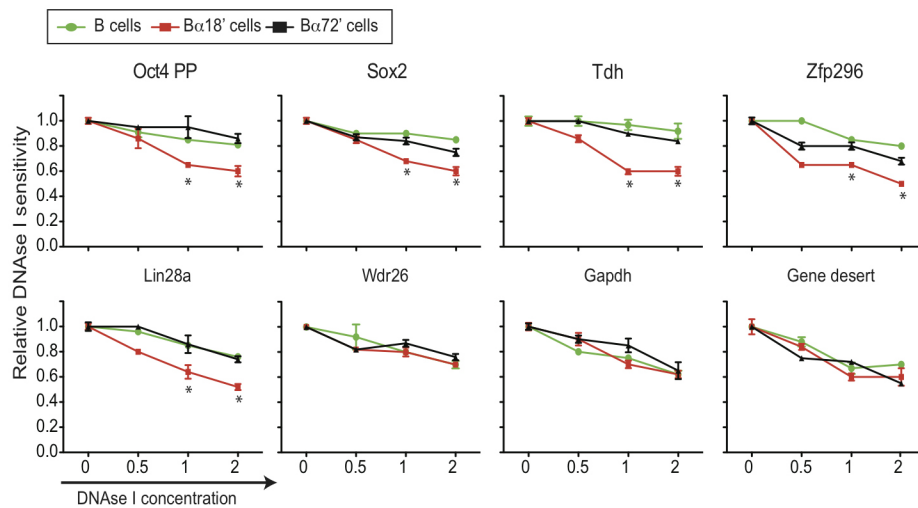
mm9 annotation. **e**, Fraction of upregulated (red chart) and downregulated genes (blue) after a pulse of C/EBP α showing a peak at 50 kb from TSS. **f**, Variation of expression after C/EBP α pulse of nearest genes to peaks. **g**, Selected gene sets significantly enriched in genes adjacent to C/EBP α peaks based on hypergeometric tests (from MSigDB cp.v4 database). Bonferroni procedure, P value <0.05 .



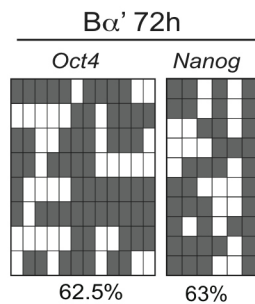
Extended Data Figure 8 | Tet2 expression in B cells and MEFs. **a**, RT-qPCR of Tet2, pluripotency and MET genes after induction of OSKM in MEFs pulsed or not with C/EBP α . Error bars indicate s.d. ($n = 3$). **b**, **c**, hMeDIP and Tet2 binding at regulatory regions of the Tet2 target gene *Hal1* (ref. 3) and *Sall4*.

d, Tet2 localization in B cells pulsed with C/EBP α for 72 h. **e**, Oestradiol treatment has no effect on cellular localization of Tet2 protein in B cells (scale bar, 30 μ m). **f**, Cellular localization of Tet2 protein in MEFs treated or not with E2 (scale bar, 100 μ m).

a



b

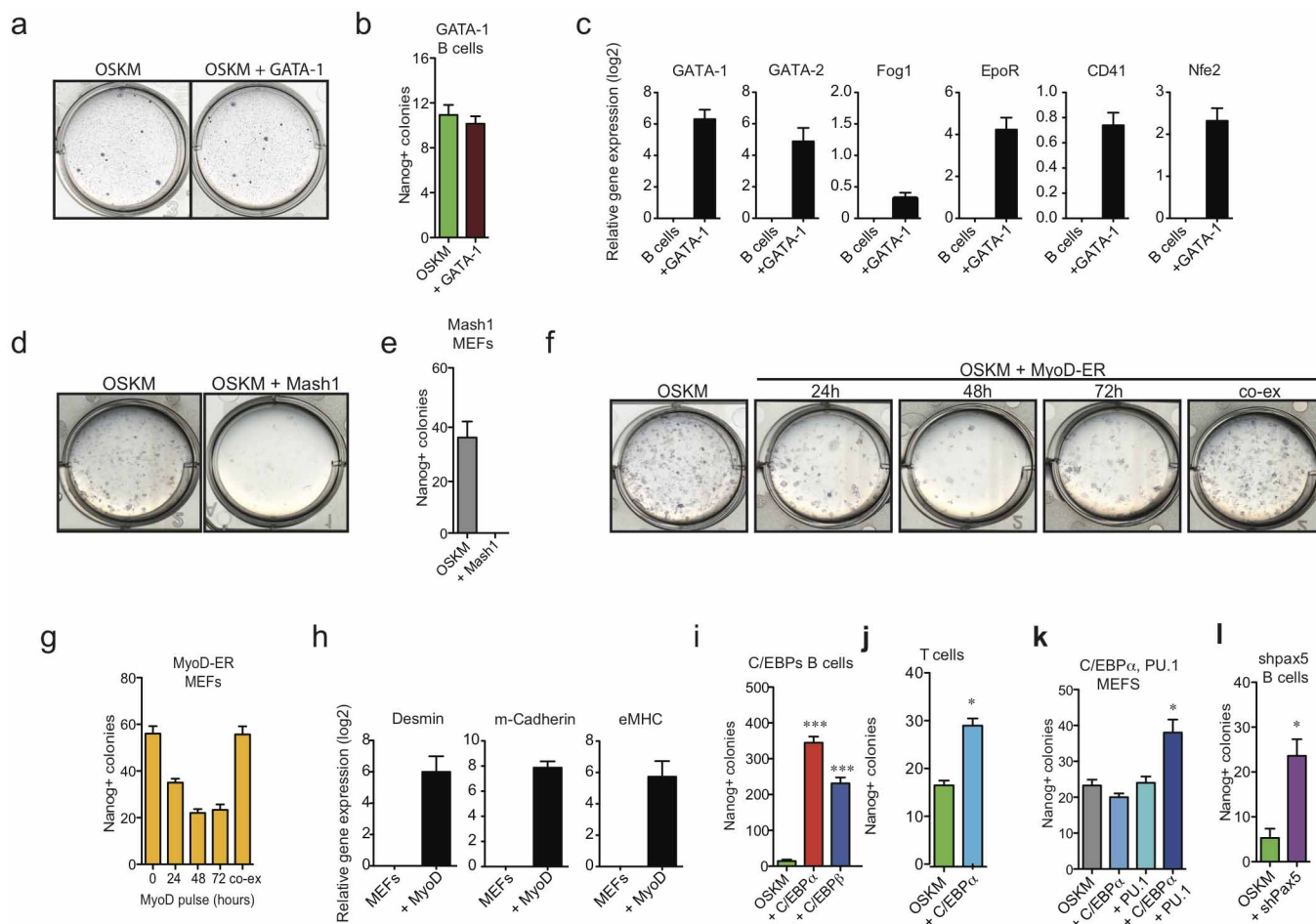


c

	B	Bα'	Bα' 2d	Bα' 4d	Bα' 6d	Bα' 8d	αiPS	Genes
1.wave PP genes								<i>Oct4, Lin28, Tdh, Gdf3, Zfp296</i>
2.wave PP genes								<i>Nanog, Sox2, Esrrb, Sall4, Zfp42, Dppa5a</i>
'Very late' PP genes								<i>Morc1, Zbtb32, Klf5, Tex19.1, Rhox9</i>
Epithelial genes								<i>Cdh1, Cdn3, Cdn4, Cdn7, Epcam</i>
Mesenchymal genes								<i>Snai1, 2, Foxc2, Col1a2, Col7a1, Twist1, 2</i>
TGFβ pathway								<i>Tgfb2, Tgfb1, Tgfb3, Tgfb4, Smad6</i>
MΦ assoc. genes								<i>Cebpa, Cebpb, Nab2</i>
B cell assoc. genes								<i>Pax5, Ebf1, Tcf3, Bcl11a</i>

Extended Data Figure 9 | DNase I sensitivity and gene expression summary. a, DNase I sensitivity at regulatory regions in B cells and Bα' cells (data from Fig. 4a) and B cells pulsed for 72 h determined by enzyme titration. Error bars indicate s.d. ($n = 3$). Student's t -test $*P < 0.05$ relative to control. b, CpGs

analysed for methylation by bisulphite sequencing in the *Pou5f1* and *Nanog* promoters. Filled rectangles represent methylated CpGs, empty rectangles unmethylated residues. c, Summary of relevant gene expression changes during the transition from B cells to Bα' cells and Bα' cells to αiPS cells.



Extended Data Figure 10 | Transcription factor and cell type specificity.

a, b, B+OSKM cells were infected with a retrovirus expressing GATA1 or left uninfected, induced with doxycycline and scored for 12 d.p.i. alkaline phosphatase⁺ and Nanog⁺ colonies. Error bars indicate s.d. ($n = 3$). **c**, qRT-PCR for megakaryocyte/erythroid-restricted genes in B cells infected with GATA1 retrovirus for 3 days. Error bars indicate s.d. ($n = 3$). **d, e**, Alkaline phosphatase⁺ colonies at 15 d.p.i. in OSKM MEFs infected with a retrovirus expressing Mash1 or mock infected, and Nanog⁺ colonies counted. Error bars indicate s.d. ($n = 3$). **f, g**, Effect of MyoD expression on iPS cell reprogramming of MEFs. Cells were infected with MyoD-ER and sequentially induced with E2 to activate MyoD and doxycycline to activate OSKM. Representative plates

with alkaline phosphatase⁺ iPS cell colonies 15 d.p.i. of MEFs pre-induced for 24, 48 or 72 h or continuously with E2. Error bars indicate s.d. ($n = 3$).

h, Expression of muscle-restricted genes in OSKM MEFs 3 days after MyoD-ER induction. Error bars indicate s.d. ($n = 3$). **i**, Effects of C/EBPα and C/EBPβ in B+OSKM cells. Student's t -test *** $P < 0.001$ relative to control. Error bars indicate s.d. ($n = 3$). **j**, Effect of an 18-h C/EBPα pulse in pre-T+OSKM cells on the formation of 12 d.p.i. Nanog⁺ colonies. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **k**, Effects of C/EBPα, PU.1 and their combination in MEFs. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$). **l**, Effect of Pax5 knock down in B-cell reprogramming. Student's t -test * $P < 0.05$ relative to control. Error bars indicate s.d. ($n = 3$).

Leukaemogenesis induced by an activating β -catenin mutation in osteoblasts

Aruna Kode¹, John S. Manavalan¹, Ioanna Mosialou¹, Govind Bhagat², Chozha V. Rathinam³, Na Luo¹, Hossein Khiabani⁴, Albert Lee⁴, Vundavalli V. Murty⁵, Richard Friedman⁶, Andrea Brum^{1,7}, David Park⁸, Naomi Galili⁹, Siddhartha Mukherjee¹⁰, Julie Teruya-Feldstein⁸, Azra Raza⁹, Raul Rabadan⁴, Ellin Berman¹¹ & Stavroula Kousteni^{1,12}

Cells of the osteoblast lineage affect the homing^{1,2} and the number of long-term repopulating haematopoietic stem cells^{3,4}, haematopoietic stem cell mobilization and lineage determination and B cell lymphopoiesis⁵⁻⁷. Osteoblasts were recently implicated in pre-leukaemic conditions in mice^{8,9}. However, a single genetic change in osteoblasts that can induce leukaemogenesis has not been shown. Here we show that an activating mutation of β -catenin in mouse osteoblasts alters the differentiation potential of myeloid and lymphoid progenitors leading to development of acute myeloid leukaemia with common chromosomal aberrations and cell autonomous progression. Activated β -catenin stimulates expression of the Notch ligand jagged 1 in osteoblasts. Subsequent activation of Notch signalling in haematopoietic stem cell progenitors induces the malignant changes. Genetic or pharmacological inhibition of Notch signalling ameliorates acute myeloid leukaemia and demonstrates the pathogenic role of the Notch pathway. In 38% of patients with myelodysplastic syndromes or acute myeloid leukaemia, increased β -catenin signalling and nuclear accumulation was identified in osteoblasts and these patients showed increased Notch signalling in haematopoietic cells. These findings demonstrate that genetic alterations in osteoblasts can induce acute myeloid leukaemia, identify molecular signals leading to this transformation and suggest a potential novel pharmacotherapeutic approach to acute myeloid leukaemia.

Mice expressing a constitutively active β -catenin allele in osteoblasts, referred to here as *Ctnnb1*^{CAosb} (CA, constitutively active; osb, osteoblast specific constitutive activity) are osteopetrotic¹⁰, and die before 6 weeks of age (Fig. 1a) for unknown reasons. Upon further examination, *Ctnnb1*^{CAosb} mice were anaemic at 2 weeks of age with peripheral blood monocytosis, neutrophilia, lymphocytopenia and thrombocytopenia (Extended Data Fig. 1a). Erythroid cells were decreased in the marrow and extramedullary haematopoiesis was observed in the liver (Fig. 1c and Extended Data Fig. 1b, l, m). Although the number of myeloid (CD11b⁺/Gr1⁺) cells decreased due to osteopetrosis, their relative percentage increased, indicating a shift in the differentiation of HSCs to the myeloid lineage (Fig. 1d and Extended Data Fig. 1c, d). The haematopoietic stem and progenitor cell (HSPC) population in the bone marrow (Lin⁻Sca⁺c-Kit⁺, LSK) cells decreased twofold in *Ctnnb1*^{CAosb} mice, but their percentage was twofold greater than in wild-type littermates (Fig. 1e and Extended Data Fig. 1e, f). The long-term repopulating HSC progenitors (LT-HSCs) increased in numbers and percentage, whereas the lymphoid-biased multipotential progenitors, LSK⁺/FLT3⁺, and the granulocyte/monocyte progenitors (GMP) (Extended Data Fig. 1g-j) decreased. The GMP percentage increased (Fig. 1f). Identical abnormalities were observed in the spleen of *Ctnnb1*^{CAosb}

mice (Extended Data Fig. 1n-p). The mutation was introduced in osteoblasts but not in any cells of the haematopoietic compartment (Extended Data Fig. 1q-t) of *Ctnnb1*^{CAosb} mice.

Blasts (12–90%) and dysplastic neutrophils (13–81%) were noted in the blood and there was dense and diffuse infiltration with myeloid and monocytic cells, blasts (30–53% for $n = 12$ mice) and dysplastic neutrophils in the marrow and spleen of *Ctnnb1*^{CAosb} mice (Fig. 1g–k, Extended Data Fig. 2a–c). In the liver, clusters of immature cells with atypical nuclear appearance were seen (Fig. 1l). The increase in immature myeloid cells was confirmed by staining with myeloid markers in bones, spleen and liver (Extended Data Fig. 2d–h). Reduced B-cell lymphopoiesis without changes in T-cell populations was observed in *Ctnnb1*^{CAosb} mice (Extended Data Fig. 2i–t). Differentiation blockade was demonstrated by the presence of immature myeloid progenitors in *Ctnnb1*^{CAosb} marrow and differentiation cultures (Fig. 1m, n and Extended Data Fig. 2u–x). These cellular abnormalities fulfil the criteria of AML diagnosis in mice¹¹ with principle features of human AML^{12,13}.

A clonal abnormality involving a Robertsonian translocation Rb(1;19) was identified in myeloid cells of the spleen of a *Ctnnb1*^{CAosb} mouse (Extended Data Fig. 2y). Recurrent numerical and structural chromosomal alterations were also detected in myeloid cells of the spleen of all mutant mice examined (Fig. 2a and Extended Data Table 1). Frequent abnormalities were detected in chromosome 5, the mouse orthologue of human chromosome 7q associated with common cytogenetic abnormalities in patients with myelodysplastic syndromes (MDS) or acute myeloid leukaemia (AML)¹⁴. Whole-exome sequencing identified 4 non-silent somatic mutations in myeloid cells from 3 *Ctnnb1*^{CAosb} mice (Fig. 2b and Extended Data Fig. 2z), including a recurrent one in *Tnfrsf21* and a single somatic mutation in *Crb1* previously reported in human AML¹⁵, but sample size has insufficient statistical power to determine if it is a driver or passenger mutation. Hence, constitutive activation of β -catenin in osteoblasts facilitates clonal progression and is associated with somatic mutations in myeloid progenitors.

Transplantation of bone marrow cells from *Ctnnb1*^{CAosb} leukaemic mice into lethally irradiated wild-type recipients induced all features of haematopoietic dysfunction and AML observed in *Ctnnb1*^{CAosb} mice including blasts (15–80%) and dysplastic neutrophils (15–75%) in the blood and blasts (30–40%) and abnormal megakaryocytes in the marrow and early lethality (Extended Data Fig. 3a–i). Transplantation of wild-type bone marrow cells to lethally irradiated *Ctnnb1*^{CAosb} mice also resulted in AML with early lethality (Extended Data Fig. 3j–r). Transplantation of LT-HSCs, but not other haematopoietic populations, from *Ctnnb1*^{CAosb} mice to sub-lethally irradiated wild-type recipients resulted in AML with early lethality (Fig. 2c, d and Extended

¹Department of Medicine, Division of Endocrinology, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA. ²Department of Pathology and Cell Biology, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA. ³Department of Genetics and Development College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA. ⁴Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, USA. ⁵Department of Pathology & Institute for Cancer Genetics Irving Cancer Research Center, Columbia University, New York, New York 10032, USA. ⁶Biomedical Informatics Shared Resource, Herbert Irving Comprehensive Cancer Center and Department of Biomedical Informatics, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA. ⁷Department of Internal Medicine, Erasmus MC, Dr. Molewaterplein 50, NL-3015 GE Rotterdam, The Netherlands. ⁸Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. ⁹Myelodysplastic Syndromes Center, Columbia University New York, New York 10032, USA. ¹⁰Departments of Medicine Hematology & Oncology Columbia University New York, New York 10032, USA. ¹¹Leukemia Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. ¹²Department of Physiology & Cellular Biophysics, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA.

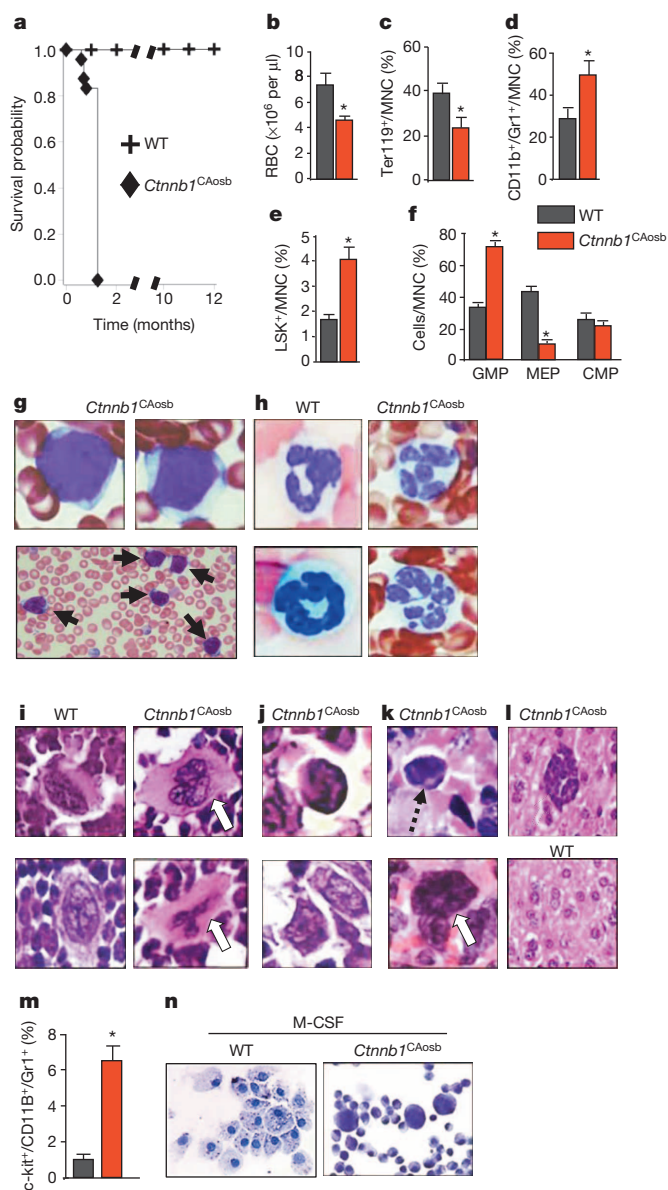


Figure 1 | Anaemia and myeloid lineage expansion in *Ctnnb1*^{CAosb} mice. **a**, Lethality. **b**, Anaemia. **c–h**, Decreased erythroid progenitors and increased percentage of monocytic/granulocytic cells (**d**), LSK cells (**e**), myeloid progenitor populations in the marrow (**f**), immature monocytic blasts (**g**) and hypersegmented neutrophils in the blood (**h**) (13–81% neutrophils and 12–90% blasts). **i, j**, Bone marrow sections showing micro-megakaryocytes with hyperchromatic nuclei (**i**) and blasts (**j**). **k**, Spleen cells with large nucleoli (dotted arrow) and dysplastic megakaryocytes (white arrow). **l**, Cluster of immature cells with atypical nuclear appearance in the liver. **m**, Increased percentage of undifferentiated immature myeloid cells in the bone marrow of *Ctnnb1*^{CAosb} mice. **n**, Lack of myeloid cell differentiation in *Ctnnb1*^{CAosb} bone marrow cells. $n = 8$ mice per wild-type and 12 mice per *Ctnnb1*^{CAosb} groups. M-CSF, macrophage colony-stimulating factor. Results show a representative of five independent experiments, * $P < 0.05$ versus wild type. Results show mean \pm s.d.; MNC, mononuclear cells.

Data Fig. 3s–z) indicating that LT-HSCs are the leukaemia-initiating cells (LICs). These results demonstrate that osteoblasts are the cells responsible for AML development in this model. Remarkably, HSCs of *Ctnnb1*^{CAosb} mice have acquired a permanent self-perpetuating genetic alteration that becomes independent of the initial mutation in osteoblasts.

All *Ctnnb1*^{CAosb} mice examined develop AML between 2 (40%) and 3.5 (60%) weeks of age. Livers of *Ctnnb1*^{CAosb} newborn mice show increased LSK cells and cells of the myeloid lineage, and a decrease in

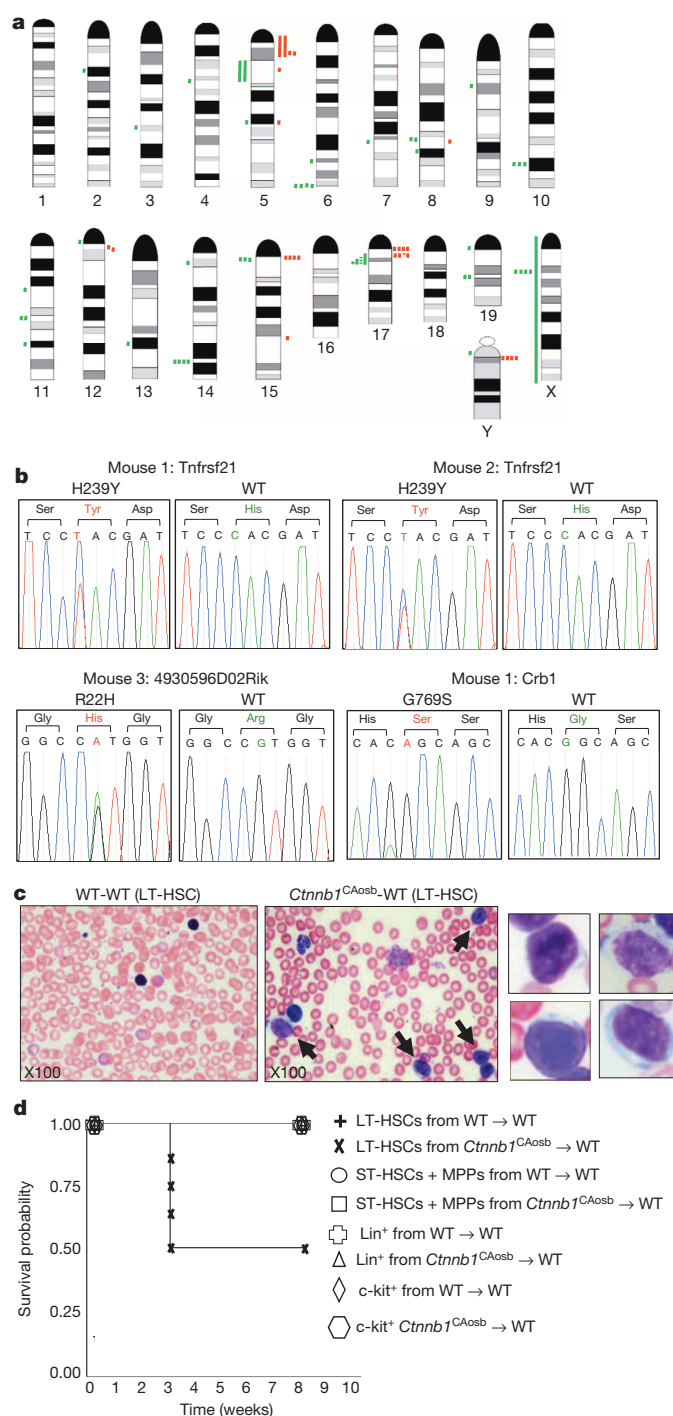


Figure 2 | AML in *Ctnnb1*^{CAosb} mice. **a**, Mouse chromosomal ideogram showing areas of genetic gain (red bars) and loss (green bars) identified by aCGH in *Ctnnb1*^{CAosb} mice. $n = 5$. **b**, Sequence traces of somatic mutations in myeloid malignancies (CD11b⁺/Gr1⁺) from 3 *Ctnnb1*^{CAosb} mice. **c**, Blasts (12–75%, solid arrows) in blood of sub-lethally irradiated CD45.1 wild-type mice transplanted with LT-HSCs from CD45.2 *Ctnnb1*^{CAosb} mice, 4 weeks following transplantation. **d**, Lethality in wild-type mice transplanted with indicated haematopoietic populations from *Ctnnb1*^{CAosb} mice. $n = 7$. Results are representative of two independent experiments.

erythroid and B-lymphoid cells (Extended Data Fig. 4a–j). Microhypolobated megakaryocytes, Pelger Huet neutrophils, seen in MDS and other congenital entities, and nuclear cytoplasmic asynchrony in the erythroid lineage were also seen in the liver and bone marrow of newborn *Ctnnb1*^{CAosb} mice, while their spleens showed increased number of blasts and a shift towards the myeloid lineage (Extended Data Fig. 4k–m).

These characteristics indicate deregulated haematopoiesis with neutrophil dyspoiesis at birth. Less than 20% blasts were seen in the marrow, consistent with a diagnosis of MDS with excess blasts (RAEB1/2). Differentiation blockade was not observed in newborn animals and fetal HSCs did not transfer the disease (Extended Data Fig. 4n–w) due to lack of HSC–osteoblast interaction in the fetal liver. These results confirm that AML is induced by defective niche signals that are restricted to the bone marrow osteoblasts.

β -catenin target genes in osteoblasts that may regulate HSC fate were identified by microarray analysis. One gene, the Notch ligand jagged 1, fulfilled 4 criteria: (1) acts on adjacent cells; (2) activates a pathway of which many targets are increased in the array; (3) has been implicated in haematopoiesis; and (4) is regulated transcriptionally by β -catenin (Extended Data Fig. 5a–d and ref. 16). Accordingly, jagged 1 expression was increased in *Ctnnb1*^{CAosb} bones and expression of the Notch targets *Hes1*, *Hes5*, *Hey1*, *Hey2* increased and *Hes1* targets *Cebpa* (also known as *Cebp α*) and *Spi1* (also known as *Pu.1*) decreased in *Ctnnb1*^{CAosb} LSK cells of *Ctnnb1*^{CAosb} mice, indicating increased Notch signalling in this population (Fig. 3a, b and Extended Data Fig. 5a, b, f, g). Notch1 and Notch2 expression was not affected (Extended Data Fig. 5e). Increased Notch signalling occurred specifically in the leukaemia-initiating LT-HSCs without changes in the other LSK compartments (Extended Data Fig. 5f, g).

To determine if jagged 1 in osteoblasts contributes to AML development in *Ctnnb1*^{CAosb} mice, we removed one allele of jagged 1 in osteoblasts (*Ctnnb1*^{CAosb}; *Jag1*^{osb}^{+/-} mice). These genetic manipulations decreased Notch signalling in LSK cells, rescued anaemia and deregulation of HSC lineage differentiation and prevented AML development (Fig. 3d–f, Extended Data Fig. 6a–j). *Ctnnb1*^{CAosb}; *Jag1*^{osb}^{+/-} mice survived and were healthy for the entire time they were observed, even though they remained osteopetrotic (Fig. 3g and Extended Data Fig. 6k). Similarly, pharmacological inhibition of Notch signalling with a γ -secretase inhibitor¹⁷ reversed haematopoietic deregulation and myeloid expansion in blood, marrow and spleen and reversed AML in *Ctnnb1*^{CAosb} mice without affecting osteopetrosis (Extended Data Figs 5h–s and 7), indicating that osteopetrosis is not enough to drive AML. These observations suggest that Notch signalling is required for AML development in *Ctnnb1*^{CAosb} mice and that chromosomal alterations may result from increased Notch signalling¹⁸. Alternatively, healthy HSCs in the endothelial and perivascular niche can multiply and outgrow leukaemic HSCs in DBZ-treated *Ctnnb1*^{CAosb} mice. The gene jagged 1 is required for leukaemia induction; whether it is involved in leukaemia maintenance with a therapeutic benefit remains to be examined.

To assess the relevance of these findings to humans we examined activation of β -catenin signalling in bone marrow biopsies from MDS or AML patients. Forty-one out of 107 patients examined with all MDS subtypes, AML, or MDS that had transformed to AML (38.3%) showed nuclear localization of β -catenin in osteoblasts (Fig. 4a, b, Extended Data Figs 8a–h and 9h and Supplementary Table 1) but in none of the 56 healthy controls examined (Fig. 4c and Extended Data Fig. 9a–g, i, j). Myeloid and erythroid cells and megakaryocytes in all patients and healthy control subjects showed membrane staining for β -catenin. Notch signalling was specifically activated only in patients with nuclear accumulation of β -catenin as indicated by HEY1 nuclear staining in their haematopoietic cells (Fig. 4d and Extended Data Fig. 8a–f). Expression of all examined β -catenin target genes and *JAG1* and *DLL1* was upregulated over twofold in osteoblasts from MDS/AML patients with β -catenin nuclear accumulation in osteoblasts (Fig. 4h, i) but not in healthy controls. Notch activity was increased in haematopoietic cells from the same patients, but not healthy controls, as indicated by two-fold increase in the expression of Notch transcriptional targets (Fig. 4j). It is possible that aberrant β -catenin signalling in osteoblasts of these patients may be the consequence of haematopoietic clones altering expansion or functionality of different stromal cell lineages, as recently reported¹⁹. During screening of assumed healthy controls, two individuals had nuclear β -catenin in osteoblasts. Re-evaluation showed that one patient

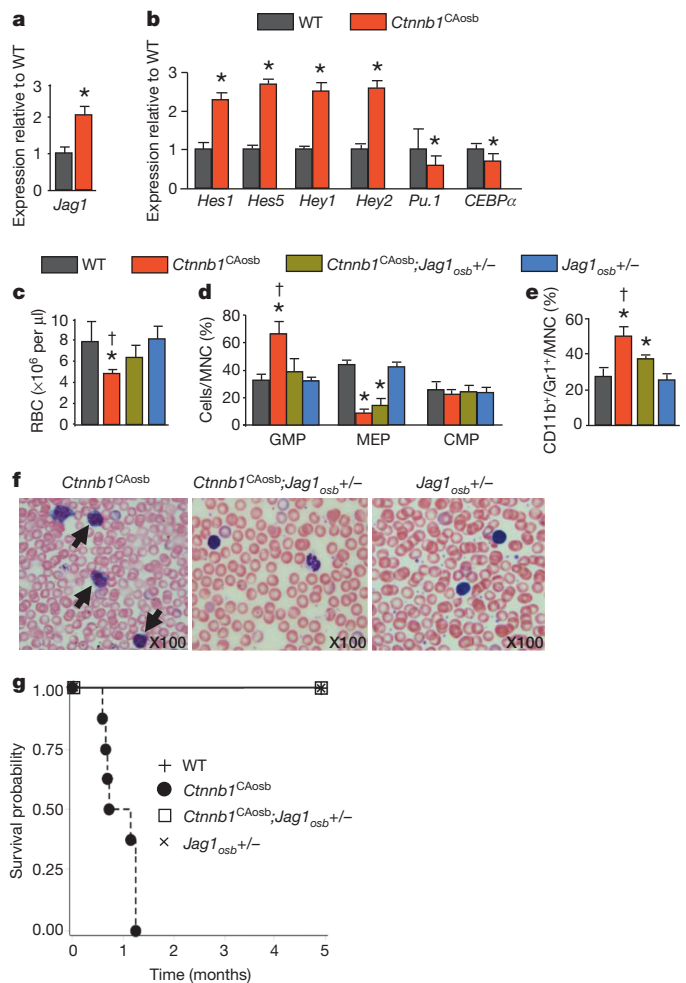


Figure 3 | Inactivation of jagged 1 in osteoblasts prevents AML in *Ctnnb1*^{CAosb} mice. **a, b**, Expression of jagged 1 in bone (a) and Notch targets in LSK⁺ cells (b) ($n = 4$). **c–e**, The bone marrow of *Ctnnb1*^{CAosb}; *Jag1*^{osb}^{+/-} mice shows rescue of anaemia (c) and of proportions of myeloid (d) and CD11b⁺/Gr1⁺ (e) cells. **f**, Blood histology. **g**, Survival of *Ctnnb1*^{CAosb}; *Jag1*^{osb}^{+/-} mice. * $P < 0.05$ versus wild type and † $P < 0.05$ *Ctnnb1*^{CAosb} versus *Ctnnb1*^{CAosb}; *Jag1*^{osb}^{+/-} mice. $n = 8$. Results show a representative of two independent experiments. Results are mean \pm s.d.

developed MDS and the second an underlying MPN/MDS, a pre-AML condition with features of both a myeloproliferative neoplasm (MPN) and MDS (Extended Data Fig. 8g, h) suggesting a potential prognostic value.

Notch activation promotes expansion of myeloid cells²⁰ and acute megakaryoblastic leukaemia-like disease in mice²¹. Other studies show that the Notch pathway may act as tumour suppressor in AML^{22–24}. However, in these models, LICs are found in GMPs, whereas in our model LICs are in LT-HSCs, suggesting that different LICs can have distinct consequences. Additionally, increased mammalian jagged 1 expression may not elicit identical outcomes as increased Notch signalling by all Notch receptors^{25–27} and *Ctnnb1*^{CAosb} osteoblasts may stimulate additional signals that act in combination with Notch to induce mutations contributing to AML. Notch also has a role in T-cell acute lymphoblastic leukaemia (T-ALL) pathogenesis²⁸, but T-cell-specific cooperative signals seem to be required to induce transformation²⁹.

The idea that osteolineage cells can induce myeloid malignancies has been previously put forward⁹. Our observations that osteoblasts determine the appearance of cell-autonomous AML with 100% penetrance, and the molecular and genetic dissection of how this occurs in mice and humans, demonstrate the role of the marrow niche as a determinant of haematological disorders. This finding may also be informative

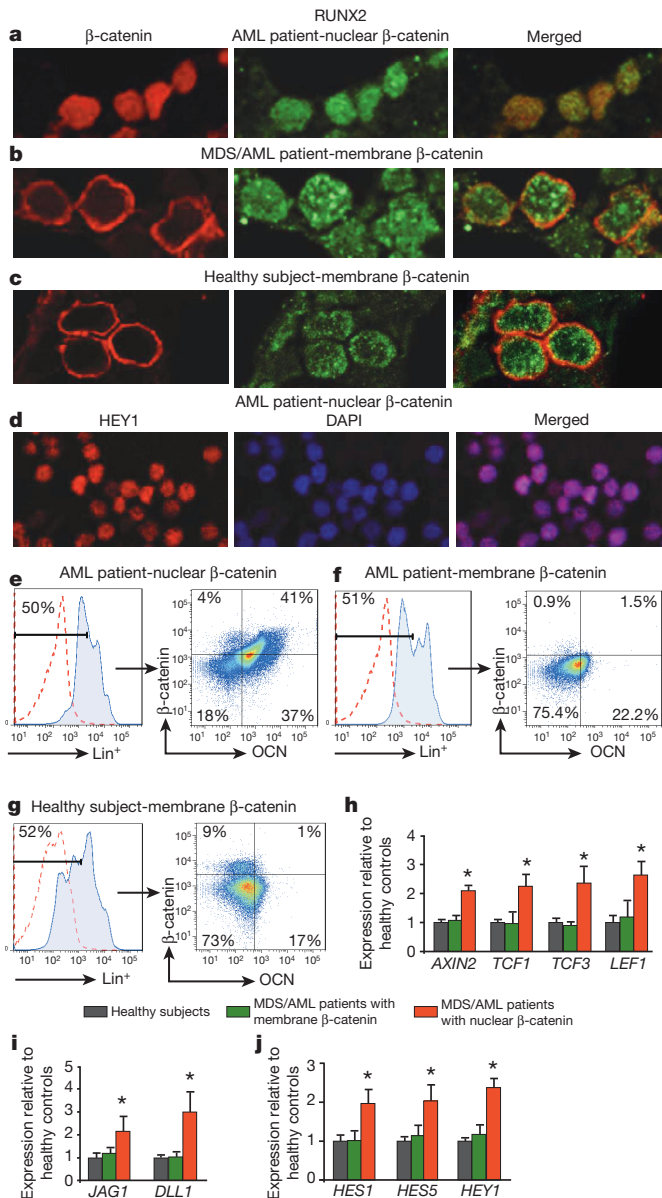


Figure 4 | Nuclear accumulation of β -catenin in osteoblasts and increased Notch signalling in human MDS/AML. **a–c**, Double immunofluorescence staining for β -catenin and RUNX2 in bone marrow biopsies from MDS/AML patients (**a**, **b**) and healthy subjects (**c**). **d**, Nuclear accumulation of HEY1 in patient shown in (**a**). **e–g**, Flow cytometry detecting nuclear or activated β -catenin. Plots show nuclear (**e**), versus non-nuclear (**f**), localization of β -catenin in osteoblasts from MDS/AML patients or $CD34^{-}/Lin^{-}OCN^{+}$ cells (OCN, osteocalcin) in healthy subjects (**g**). In **a–g**, one representative data piece from 107 patients and from 56 healthy subjects. **h–j**, Samples from MDS/AML patients or healthy subjects show increased expression of β -catenin targets (**h**), and *JAG1* and *DLL1* (**i**) in $CD34^{-}/Lin^{-}OCN^{+}$ osteoblasts and Notch (**j**) targets in $CD34^{+}/Lin^{+}OCN^{-}$ haematopoietic cells. $P < 0.05$ versus patients with non-nuclear β -catenin in osteoblasts and healthy subjects. Results are mean \pm s.d. Results show a representative of two independent experiments with $n = 3$ for healthy subjects, 12 for MDS/AML patients with membrane localization of β -catenin and 11 for MDS/AML patients with nuclear β -catenin.

about MDS/AML pathogenesis in humans and expand the potential of new therapeutic applications.

METHODS SUMMARY

Mice. Generation of $\alpha 1(I)Collagen-Cre$, $Catnb^{+/lox(ex3)}$, $Ctnnb1^{CAosb}$ and $Jag1^{fl/fl}$ mice have previously been reported. The original nomenclature used for these

mice was $\beta cat(ex3)osb$. $Catnb^{+/lox(ex3)}$ mice express a β -catenin mutant allele in which exon 3, encoding all serine and threonine residues phosphorylated by glycogen synthase kinase 3 β , is flanked by *loxP* sites. All the protocols and experiments were conducted according to the guidelines of the Institute of Comparative Medicine, Columbia University.

Patient samples. Bone marrow biopsies from patients with AML and MDS were consecutively obtained from 2000 to 2008 and reviewed under a research exempt waiver approved by the institutional review boards (IRB) of Memorial Sloan-Kettering Hospital and Columbia University and Human Biospecimen Utilization Committee.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February; accepted 14 November 2013.

Published online 15 January 2014.

- Heissig, B. *et al.* Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of Kit-ligand. *Cell* **109**, 625–637 (2002).
- Shiozawa, Y. *et al.* Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J. Clin. Invest.* **121**, 1298–1312 (2011).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Wu, J. Y. *et al.* Osteoblastic regulation of B lymphopoiesis is mediated by Gs α -dependent signaling pathways. *Proc. Natl Acad. Sci. USA* **105**, 16976–16981 (2008).
- Zhu, J. *et al.* Osteoblasts support B-lymphocyte commitment and differentiation from hematopoietic stem cells. *Blood* **109**, 3706–3712 (2007).
- Chan, C. K. *et al.* Endochondral ossification is required for haematopoietic stem-cell niche formation. *Nature* **457**, 490–494 (2009).
- Walkey, C. R. *et al.* A microenvironment-induced myeloproliferative syndrome caused by retinoic acid receptor gamma deficiency. *Cell* **129**, 1097–1110 (2007).
- Raaijmakers, M. H. *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* **464**, 852–857 (2010).
- Glass, D. A. *et al.* Canonical Wnt signaling in differentiated osteoblasts controls osteoclast differentiation. *Dev. Cell* **8**, 751–764 (2005).
- Kogan, S. C. *et al.* Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood* **100**, 238–245 (2002).
- Sternberg, A. *et al.* Evidence for reduced B-cell progenitors in early (low-risk) myelodysplastic syndrome. *Blood* **106**, 2982–2991 (2005).
- van de Loosdrecht, A. A. *et al.* Identification of distinct prognostic subgroups in low- and intermediate-1-risk myelodysplastic syndromes by flow cytometry. *Blood* **111**, 1067–1077 (2008).
- Raza, A. & Galili, N. The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. *Nature Rev. Cancer* **12**, 849–859 (2012).
- The Cancer Genome Atlas Research Network Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
- Estrach, S., Ambler, C. A., Lo, C. C., Hozumi, K. & Watt, F. M. Jagged 1 is a β -catenin target gene required for ectopic hair follicle formation in adult epidermis. *Development* **133**, 4427–4438 (2006).
- Real, P. J. *et al.* γ -secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia. *Nature Med.* **15**, 50–58 (2009).
- Baia, G. S. *et al.* Notch activation is associated with tetraploidy and enhanced chromosomal instability in meningiomas. *Neoplasia* **10**, 604–612 (2008).
- Schepers, K. *et al.* Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. *Cell Stem Cell* **13**, 285–299 (2013).
- Delaney, C. *et al.* Notch-mediated expansion of human cord blood progenitor cells capable of rapid myeloid reconstitution. *Nature Med.* **16**, 232–236 (2010).
- Mercher, T. *et al.* The OTT-MAL fusion oncogene activates RBPJ-mediated transcription and induces acute megakaryoblastic leukemia in a knockin mouse model. *J. Clin. Invest.* **119**, 852–864 (2009).
- Lobry, C. *et al.* Notch pathway activation targets AML-initiating cell homeostasis and differentiation. *J. Exp. Med.* **210**, 301–319 (2013).
- Klinakis, A. *et al.* A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature* **473**, 230–233 (2011).
- Kannan, S. *et al.* Notch activation inhibits AML growth and survival: a potential therapeutic approach. *J. Exp. Med.* **210**, 321–337 (2013).
- Krebs, L. T. *et al.* Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev.* **14**, 1343–1352 (2000).
- Kumano, K. *et al.* Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. *Immunity* **18**, 699–711 (2003).
- Mancini, S. J. *et al.* Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. *Blood* **105**, 2340–2342 (2005).
- Lobry, C., Oh, P. & Aifantis, I. Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think. *J. Exp. Med.* **208**, 1931–1935 (2011).

29. Grabher, C. von Boehmer, H. & Look, A. T. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nature Rev. Cancer* **6**, 347–359 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are grateful to M. Taketo for providing the *Catnb*^{+/lox(ex3)} mice; R. Dalla-Favera and G. Karsenty for helpful discussions and critical reading of the manuscript; M. Van Den Brink for providing funding support; D. McMahon and C. A. Zhang for help with statistical analysis; and J. Sha, F. N. Emmons, I. Linkov and J. Pichard for technical assistance. The histology and metabolic unit facility of the Diabetes and Endocrinology Research Center (DERC, NIDDK DK063608-07), the Molecular Pathology facility of the Herbert Irving Cancer Center of Columbia University Medical Center provided help with histological analysis and W. Fantl and G. Nolan for provided flow cytometry conditions for the nuclear β -catenin antibody. This work was supported by the National Institutes of Health (R01 AR054447, and P01 AG032959 and R01 AR055931 to S.K.) and by the Division of Hematologic Oncology, Memorial Sloan-Kettering Cancer Center, New York (to E.B.) and by a European Union Marie Curie Fellowship INTERBONE–FP7–PEOPLE–2011–IRSES–295181 (to A.B.).

Author Contributions A.K. and S.K. initiated the study and designed the experiments. A.K., J.S.M. and S.K. analysed the data. A.K. carried out most of the experimental work with the help of J.S.M., I.M. and N.L. J.S.M. performed the flow cytometry analysis. H.K., A.L. and R.R. performed whole-exome sequencing analysis. I.M. confirmed exome mutations. N.L. performed immunofluorescence. C.V.R. reviewed and discussed haematopoiesis data and bone marrow transplantations. G.B., D.P. and J.T.-F. performed histology in mouse samples. J.T.-F. and D.P. performed histology in human samples. A.R., S.M., N.G., J.T.-F. and E.B. provided human AML and MDS samples and reviewed and discussed human bone marrow and bone biopsy data. M.V. performed G-banding karyotype analysis. R.F. analysed microarray data. A.B. assisted with mouse flow cytometry experiments. A.K. and S.K. wrote the manuscript. S.K. directed the research. All authors discussed and commented on the manuscript.

Author Information Microarray and aCGH data were deposited in Gene Expression Omnibus (accession numbers GSE43242, GSE51690) and exome sequencing data were deposited in Sequence Read Archive (accession number SRP031981). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K. (sk2836@columbia.edu).

METHODS

Animals. Generation of $\alpha 1(I)Collagen-Cre$ and $Catnb^{+/lox(ex3)}$ mice have previously been reported^{30–32}. $Catnb^{+/lox(ex3)}$ mice express a β -catenin mutant allele in which exon 3, encoding all serine and threonine residues phosphorylated by glycogen synthase kinase 3 β (GSK-3 β)³³, is flanked by *loxP* sites. Mice with osteoblast-specific constitutive activation of β -catenin were generated by crossing $Catnb^{+/lox(ex3)}$ mice with $\alpha 1(I)Col-Cre$ mice expressing Cre under the control of 2.3 kb of the proximal promoter of the mouse pro- $\alpha 1(I)$ collagen gene. The transgene is expressed at high levels in osteoblasts specifically³⁴. There is no expression in chondrocytes, condensed mesenchymal cells, perichondrial or periosteal fibroblasts, or any other type I collagen-producing cells, or other fibroblast-rich tissues such as muscle, heart or tendons. The resulting offspring, termed $Ctnnb1^{CAosb}$ (original nomenclature $\beta cat(ex3)_{osb}$), express a constitutively active β -catenin allele in osteoblasts. Mice with osteoblast-specific deletion of jagged 1 were generated by crossing previously described $Jag1^{fl/fl}$ mice³⁵ with $\alpha 1(I)Col-Cre$ mice. Genotyping was performed at weaning stage by PCR analysis of genomic DNA. In each experiment the mice used were all littermates of the same genetic background. One-month-old female mice were used for the bone histomorphometric analysis. $Ctnnb1^{CAosb}$ mice lack teeth due to osteopetrosis and therefore were fed a normal powdered diet that contained all the required nutrients (20% protein, 3.0 p.p.m. folic acid, 51 μ g per kg vitamin B12 from PicoLab Rodent Diet 20). Folate and vitamin B12 levels in their blood were normal (folate >24 ng ml⁻¹ and B12 >1,000 pg ml⁻¹) confirming adequate intake of critical nutrients. Folate and vitamin B12 levels were measured by Antech Diagnostics using a chemiluminescence based kit (Siemens). All the protocols and experiments were conducted according to the guidelines of the Institute of Comparative Medicine, Columbia University. Randomization was done according to genotype and blinding was applied during histological analysis.

Patient samples. Bone marrow biopsies from patients with AML and MDS were consecutively obtained from 2000 to 2008 and reviewed under a research exempt waiver approved by the institutional review board (IRB) of Memorial Sloan-Kettering Hospital and Human Biospecimen Utilization Committee. Bone marrow biopsies and aspirates obtained from Columbia University from patients with MDS and AML were stored in the IRB-approved Tissue Repository at Columbia University Medical Center after informed consent. This study was conducted under protocol approval from the IRB for use of samples from the Tissue Repository.

Karyotype analysis. Metaphase chromosome preparations were prepared from cells obtained from spleen specimens from $Ctnnb1^{CAosb}$ mice after overnight culture in complete RPMI medium using standard methods. Giemsa banding was performed and the images were captured using Cytovision Imaging system (Applied Imaging) attached to a Nikon Eclipse 600 microscope. Twenty to thirty karyotypes were prepared from each sample and described using the standard chromosome nomenclature for mice.

Array comparative genomic hybridization (aCGH). aCGH analysis was performed in the spleen of $Ctnnb1^{CAosb}$ mice using the mouse genome CGH 244A platform (Agilent Technologies) according to the manufacturer's instructions. In brief, spleen DNA from wild-type littermates was used as reference DNA. Genomic DNA was subjected to restriction digestion before labelling and purification (SureTag DNA labelling kit, Agilent Technologies). For each 244 K array, 2 μ g of labelled DNA and 2 μ g of germline reference DNA were labelled with Cy5 and Cy3, respectively. Differentially labelled test (tumour) DNA and normal reference DNA were hybridized simultaneously to normal chromosome spreads. Data extraction was conducted using the Agilent feature extraction software. Data files were analysed using the Agilent DNA analytics software. Data were deposited in Gene Expression Omnibus (accession number GSE51690).

Whole-exome capture and massively parallel sequencing, sequence mapping and identification of tumour-specific variants. For three tumour and three unpaired normal samples, purified genomic DNA (3 μ g) was enriched in protein-coding sequences using the SureSelect Mouse All Exon kit (Agilent Technologies) following standard protocols. The resulting target-enriched pool was amplified and subjected to paired-end sequencing (2 \times 100 bp) by using HiSeq2000 sequencing instruments. Exome capture and sequencing procedures were performed at Agilent Technologies. Sequencing reads were mapped to the reference genome mm10 using the Burrows-Wheeler Aligner (BWA) alignment tool version 0.5.9 (ref. 36). We identified sites that differed from the reference genome (called here variants) and constructed empirical priors for the distribution of variant frequencies in each sample independently. We obtained high-credibility intervals (posterior probability $\geq 1 - 10^{-5}$) for the observed frequency of the variants using the statistical algorithm for variant identification (SAVI) algorithm^{37,38}.

Variants were considered absent if found with a frequency between 0 and 2%, and were considered present if detected with a frequency above 15%. We chose 15% as a cut-off given its correspondence with the sensitivity threshold of direct Sanger sequencing. Variant total depth was required to be >10 \times and <300 \times . Segmenting variants that exist in one case only and absent in the other five cases

identified regions of possible copy number aberrations. We removed the variants found in these regions. We also excluded all silent variants and those present in the dbSNP database, and focused only on substitution mutations. Finally, in the tumour samples, we removed all variants found present in any of the normal samples.

The mutations were subjected to validation (present in tumour, absent in normal) by conventional Sanger-based resequencing analysis of PCR products obtained from tumour DNA using primers specific for the exon encompassing the variant. Data were deposited in the Sequence Read Archive (accession number SRP031981).

Microarray analysis. Total RNA was extracted from primary osteoblasts isolated from mouse calvaria using TRIzol reagent (Invitrogen). Microarray analysis was performed using the GeneChip 3' IVT Express kit and mouse genome 430 2.0 array gene chips (Affymetrix) according to the manufacturer's instructions. In brief, antisense RNA was synthesized from 500 ng of RNA and was biotinylated followed by purification and fragmentation using the GeneChip 3' IVT Express kit. Fragmented aRNA was hybridized to Affymetrix mouse genome 430 2.0 array gene chips. Following hybridization, chips were scanned with a GeneChip Scanner 3000 7G (Affymetrix). Data were normalized using the Mas5 method³⁹, and then log₂ transformed. Data were deposited in Gene Expression Omnibus (accession number GSE43242)⁴⁰. Differential expression was analysed using LIMMA⁴¹. We focused on about 20 genes which we selected in advance of the analysis. Genes were considered which either are active in AML, are amplified according to our CGH results, activate Notch, or whose transcription is induced by Notch. A significance cutoff of a raw $P < 0.05$ was used, as is appropriate for small previously determined gene sets⁴². Representative probe sets of genes whose expression changed greater than $\pm 20\%$ in at least one of the two mutants relative to wild type appear in Supplementary Table 1.

Bone marrow transplantation. For bone marrow transplantation studies, adult wild-type B5.SJL (CD45.1) recipient mice (8 weeks of age) were lethally irradiated (10 Gy, split dose) and were then transplanted with 1×10^5 of total bone marrow cells from $Ctnnb1^{CAosb}$ (CD45.2) or wild-type B5.SJL (CD45.2) mice (4 weeks of age) by retro-orbital venous plexus injection. Engraftment efficiency in recipients was monitored by donor contribution of CD45.2⁺ cells using FACS analysis. For the reverse experiment, because of the early lethality of $Ctnnb1^{CAosb}$ mice, 1×10^5 of total bone marrow cells from wild-type B6.SJL (CD45.1) mice were transplanted into lethally irradiated (600 rads, split dose) newborn (P1) $Ctnnb1^{CAosb}$ mice or wild-type littermates by liver injections. Engraftment efficiency in recipients was monitored by donor contribution of CD45.1⁺ cells using FACS analysis. For HSC and progenitor transplantation studies, sublethally (5.5 Gy) irradiated wild-type B5.SJL (CD45.1) recipient mice (8 weeks of age) were injected with fractionated donor bone marrow subsets isolated from $Ctnnb1^{CAosb}$ (CD45.2) or wild-type B5.SJL (CD45.2) mice (4 weeks of age). Engraftment efficiency in recipients was monitored by donor contribution of CD45.2⁺ cells using FACS analysis.

Treatment of animals with γ -secretase inhibitor. Two-week-old $Ctnnb1^{CAosb}$ mice or their wild-type littermates were treated with vehicle, the γ -secretase inhibitor DBZ ((2S)-2-[2-(3,5-difluorophenyl)-acetylaminol]-N-(5-methyl-6-oxo-6,7-dihydro-5H-dibenzo[b,d]azepin-7-yl)-propanamide, 2 μ mol kg⁻¹) daily by intraperitoneal injection for 10 days. DBZ is cell-permeable, selective, non-transition and non-competitive inhibitor of the γ -secretase complex. DBZ was synthesized to >99.9% purity as assessed by liquid chromatography/mass spectrometry (LC/MS) (Syncom) and suspended in a 0.5% methocel E4M (w/v, Colorcon) and 0.1% (v/v) Tween-80 (Sigma) solution⁴³. Immediately before intraperitoneal injection, DBZ was sonicated for 2 min to achieve a homogenous suspension.

Haematological measurements and peripheral blood morphology. For haematological measurements, blood was collected by cardiac puncture. Peripheral blood cell counts were performed on a FORCYTE haematology analyser (Oxford Science). For morphological assessment, peripheral blood smears were stained with Wright-Giemsa stain (Sigma-Aldrich) for 10 min followed by rinsing in dH₂O for 3 min. Images were taken using a 60 \times objective on a Leica microscope outfitted with a camera.

Real-time PCR. Total RNA was isolated from LSK or haematopoietic cells using RNeasy Micro Plus kit (Qiagen). Total RNA from bone marrow-free long bones was isolated using TRIzol reagent after removal of the periosteal layer. Quantitative real-time PCR was performed using the SYBR Green Master Mix (Bio-Rad) as previously described³². β -actin was used as endogenous control. Gene expression in LT-HSCs, ST-HSCs and MPPs was performed using the Power SYBR Green Cells-to-CT kit (Ambion Life Technologies).

Reporter constructs and luciferase assays. The *Jag1* promoter region carries multiple potential TCF/LEF binding sites (C/TCTTTG) located up to nucleotide -4075 (-4075, -3072, -2626, -2578, -2343, -1992, -1957, -1566, -1221, -782). The mouse reporter constructs -4112/+130 and -2100/+130 for *Jag1-luc* were generated by PCR amplification of the corresponding fragments using mouse genomic DNA as template and subsequent subcloning into the BglII and KpnI/BglII sites of the pGL3Basic vector (Promega), respectively. Transient transfection

assays were performed in HEK293T using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were seeded in 24-well plates at a density of 0.3×10^5 cells per well. Then 24 h later, cells were transfected with a total amount of 350 ng of DNA containing 150 ng reporter plasmid and 50 ng β -catenin and TCF-1 expression vectors. A total of 5 ng of pRL-CMV Renilla (Promega) was used as an internal control to normalize for transfection efficiency and equivalent amounts of DNA were achieved with pcDNA3 vector. Forty hours after transfection luciferase activity was determined using the dual luciferase reporter assay system (Promega) and quantified using Fluostar Omega (BMG Labtech). Luciferase activity is presented as fold induction over basal conditions normalized to empty luciferase vector for identical experimental conditions.

Chromatin immunoprecipitation (ChIP) assay. Primary osteoblasts were seeded in 10-cm dishes at a density of 5×10^6 . Cells were crosslinked with 1% formaldehyde for 10 min. Following Dounce homogenization, nuclei were collected and sonicated on ice to an average length of 0.5 kb. The samples were centrifuged and precleared with protein G in the presence of sonicated λ DNA and bovine serum albumin for 2 h at 4°C . One-tenth of the volume of supernatant was used as input, and the remaining volume was immunoprecipitated with β -catenin antibody and the immune complexes were collected by absorption to protein G-sepharose, washed, eluted from the beads and incubated for 5 h at 65°C to reverse crosslinks. After treatment with proteinase K, DNA was extracted with phenol-chloroform and precipitated with ethanol. Immunoprecipitated chromatin was analysed by PCR using primers corresponding to TCF/LEF binding sites on the jagged 1 promoter (−4075, −3072, −2626, −2578, −2343, −1992, −1957, −1566, −1221, −782). Putative binding elements were identified by using the TRANSFAC database. The products of the PCR amplifications were analysed by agarose gel electrophoresis.

Antibodies and flow cytometry analysis. Freshly isolated bone marrow cells and spleen cells were resuspended in flow-staining buffer (PBS plus 2% FBS) and the primary conjugated antibodies were added. After 30 min incubation at 4°C , the cells were then washed twice before flow cytometry analysis. The following monoclonal antibodies conjugated with fluorescein isothiocyanate (FITC), Allophycocyanin (APC) phycoerythrin (PE), PE-Cy7, APC-Cy7, Per-CPCy5.5, Pacific Blue, and Alexa 700 were used: CD117 (c-kit; 2B8), Sca-1 (D7), Mac-1 α (M1/70), Gr-1 (RB6-8C5), TER-119, (Ly-76) B220 (CD45R), CD19 (1D3), IgM (R6-60.2), CD3 (17A2), CD4 (RM4-5), CD8a (53-6.7), CD34 (RAM34), CD45 (30-F11), CD31 (MEC 13.3), CD16/CD32 (Fc γ RII/III; 2.4G2), CD135 (A2F10.1), CD150 (9D1), CD71 (C2), CD45.2 (104), CD45.1 (A20), F4/80, Non-phospho (active) β -catenin (S33/S37/T41) antibody, IL-7R α (SB199), Jagged-1 (C-20) (Cell Signaling; D13A1). Seven-colour flow cytometry acquisition was performed using a LSR II flow cytometer (Becton Dickinson) and analysis using FlowJo software (Treestar). Cells were gated for size, shape and granularity using forward and side scatter parameters. The positive populations were identified as cells that expressed specific levels of fluorescence activity above the nonspecific auto fluorescence of the isotype control. Nonspecific binding was reduced by preincubation with unconjugated anti-FcRII/III (2.4G2). Osteoblasts from MDS/AML patients or healthy subjects were identified as CD34 $^-$ /Lin $^-$ OCN $^+$ cells, (OCN, osteocalcin, an osteoblast-specific protein used for isolation of live osteoblastic cells). For flow sorting, bone marrow, spleen and thymus cells were resuspended in flow staining buffer at 1×10^6 cells ml $^{-1}$ and labelled with the appropriate conjugated antibodies. After 30 min incubation, cells were washed twice using flow buffer. Flow sorting was performed using FACSaria (Becton Dickinson). Sorted populations were subsequently cultured or stored in RLT buffer at -80°C for later extraction of RNA. Fluorescence intensity plots were presented in log scale. All flow cytometry data are representative of five independent experiments.

Clonogenic assay. Bone marrow cells from 4-week-old *Ctnnb1*^{CAosb} or wild-type mice were cultured in DMEM α with 10% FBS in the presence of 10 ng ml $^{-1}$ of GM-CSF or M-CSF or G-CSF for 7 days. An aliquot of the cells was used to prepare cytospins and stained with Giemsa to identify blasts. A second aliquot was analysed by flow cytometry for expression of F4/80, CD11b and Gr1.

Isolation and counting of osteoblasts from murine and human bone. The periosteal layer was removed from murine tibia and femurs, the remaining bone was crushed and washed to remove the bone marrow and bone pieces were digested with collagenase type III. Osteopetrosis in *Ctnnb1*^{CAosb} mice does not allow the use of only endosteal bone due to dispersion in the marrow space of irregular trabecular units. Human bone biopsies were dissected into pieces and fat and clot was removed from bone chips and a 3 mm section was transferred into 500 μ l α MEM with 1% penicillin/streptomycin. Scissors were used to cut the bone chip into a slurry and then the slurry was digested in 500 μ l FBS-free α MEM (1% penicillin/streptomycin) and 4 mg ml $^{-1}$ collagenase type III (Worthington) for a final concentration of 2 mg ml $^{-1}$. After incubation for 1 h with intermittent vortexing, slurry was frozen live for later use in 90% FBS with 10% DMSO. For flow cytometry analysis, osteoblasts were identified from the digested bone samples as a

population of CD34 $^-$ Lin $^-$ Ocn $^+$ cells, in which Ocn (osteocalcin) is an osteoblast-specific, non-nuclear protein commonly used for isolation of live osteoblastic cells^{44–46}. For microarray analysis and for experiments in mice, all gene expression studies were repeated using calvaria-derived cells a population rich in committed osteoblast progenitors and routinely used as osteoblast-representative. Primary murine osteoblasts were prepared from calvaria of 2-day-old pups as previously described^{47,48}. Mice calvaria were sequentially digested for 20, 40 and 90 min at 37°C in a modified minimal essential medium (Gibco); 10% FBS containing 0.1 mg of collagenase P (Worthington) per ml and 0.25% trypsin (Gibco). Cells of the first two digests were discarded, whereas cells released from the third digestion were plated in a minimal essential medium–10% FBS.

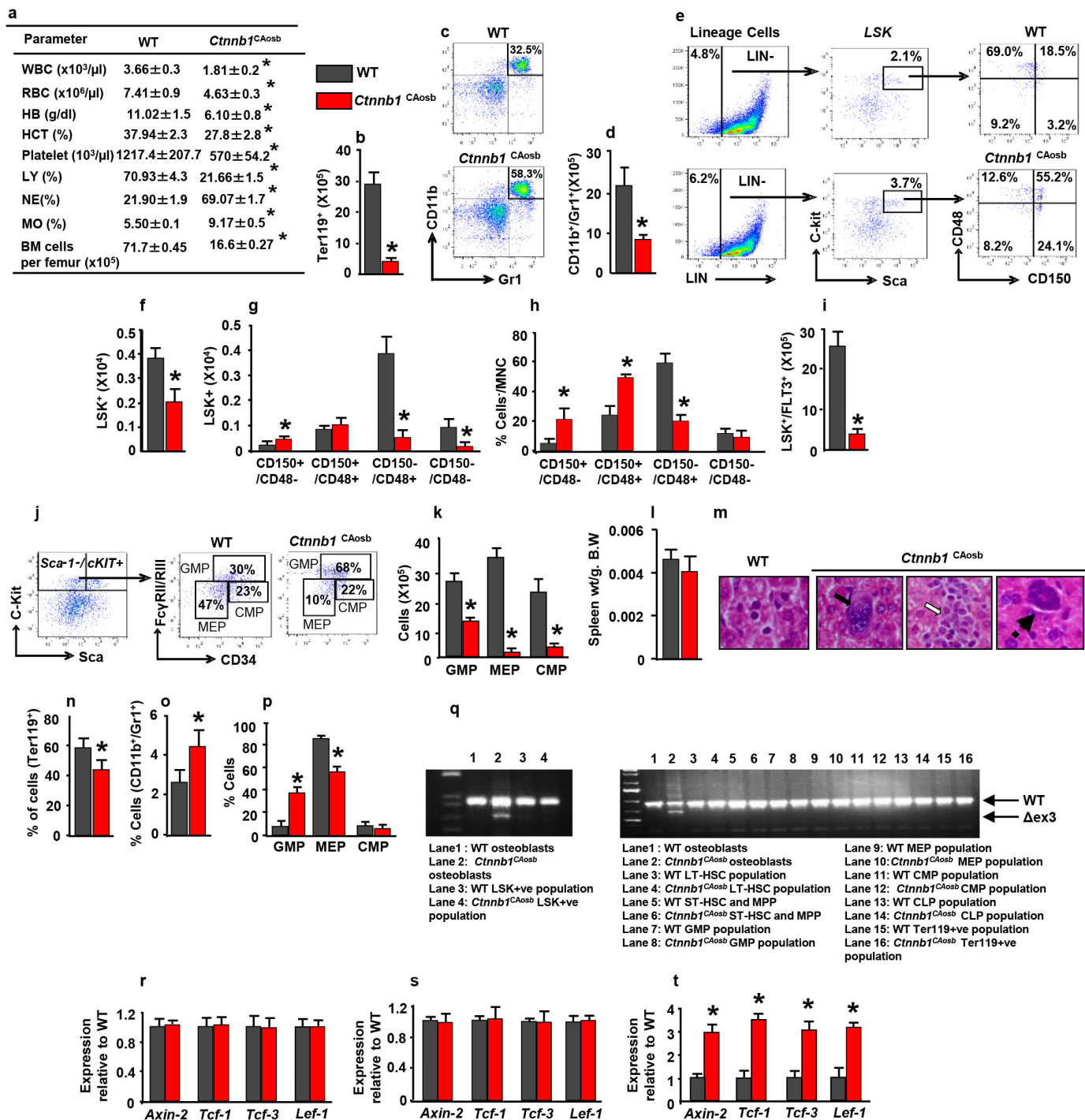
Osteoblasts were counted in each human bone biopsy as defined by standard histomorphometry guidelines^{49–51}. The number of osteoblasts per mm of bone surface was calculated. The number of osteoblasts counted depends on the size of the sample and the bio-/pathophysiological characteristics of the individual and for this study, the size of the biopsy (1 cm) allowed for counting of 30 osteoblasts per biopsy.

Histological analysis of human biopsies and murine bone, spleen and liver. Bone marrow biopsies were fixed overnight in 10% neutral formalin solution, decalcified embedded in paraffin and sectioned at 5 μ m per standard laboratory protocol. Sections were stained for β -catenin using a monoclonal antibody (1:1,000, BD Transduction Lab) or for RUNX2 using a polyclonal antibody (1:100 dilution, Santa Cruz) or for HEY1 using a polyclonal antibody (1:300 dilution, Abcam) on an automated Ventana Discovery XT platform according to the manufacturer's instructions. Immune complexes formed were developed using a DAB map kit (Ventana). Murine long bones, spleen and liver were collected from one-month-old mice, fixed overnight in 10% neutral formalin solution, embedded in paraffin, sectioned at 5 μ m and stained with haematoxylin and eosin (H&E). For immunohistochemistry, specimens were incubated with CD-117 (C-kit; Abcam), CD13 (Santa Cruz) or Myeloperoxidase (MPO) antibodies after an antigen retrieval step and blocking of endogenous peroxidase with 3% H₂O₂. Sections were then incubated with biotinylated secondary antibody and immune complexes formed were detected using standard avidin biotin complex method.

Statistical analysis. All data are represented as mean \pm standard deviation. Statistical analyses were performed using a one-way ANOVA followed by Student–Newman–Keuls test and a *P* value less than 0.05 was considered significant. Time-to-event analysis was used to assess medium survival time to death. Kaplan–Meier curves were generated to illustrate time to death, stratified by group status. Statistical significance of the between-group difference in the median time-to-endpoint was assessed by the log-rank test. Statistical analyses were performed using XLSTAT (2012.6.02, Addinsoft) and SAS (version 9.2; SAS institute). A *P* value less than 0.05 was considered statistically significant.

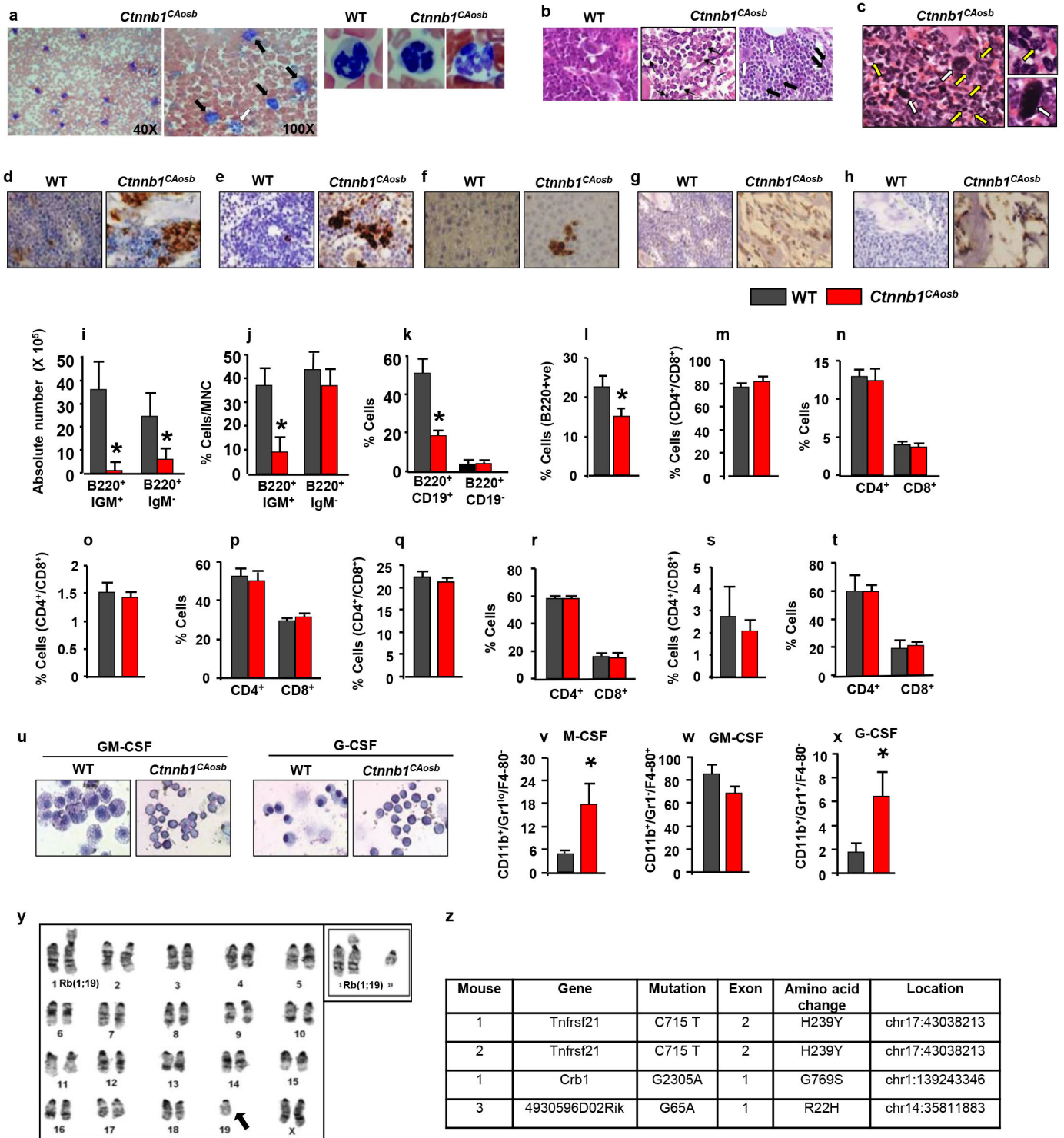
30. Dacquin, R., Starbuck, M., Schinke, T. & Karsenty, G. Mouse $\alpha 1$ (I)-collagen promoter is the best known promoter to drive efficient Cre recombinase expression in osteoblast. *Dev. Dyn.* **224**, 245–251 (2002).
31. Harada, N. *et al.* Intestinal polyposis in mice with a dominant stable mutation of the β -catenin gene. *EMBO J.* **18**, 5931–5942 (1999).
32. Rached, M. T. *et al.* FoxO1 expression in osteoblasts regulates glucose homeostasis through regulation of osteocalcin in mice. *J. Clin. Invest.* **120**, 357–368 (2010).
33. Logan, C. Y. & Nusse, R. The Wnt signalling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
34. Rossert, J., Eberspaecher, H. & de Crombrughe, B. Separate cis-acting DNA elements of the mouse pro- $\alpha 1$ (I) collagen promoter direct expression of reporter genes to different type I collagen-producing cells in transgenic mice. *J. Cell Biol.* **129**, 1421–1432 (1995).
35. Kiernan, A. E., Xu, J. & Gridley, T. The Notch ligand JAG1 is required for sensory progenitor development in the mammalian inner ear. *PLoS Genet.* **2**, e4 (2006).
36. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
37. Tiacci, E. *et al.* BRAF mutations in hairy-cell leukemia. *N. Engl. J. Med.* **364**, 2305–2315 (2011).
38. Trifonov, V., Pasqualucci, L., Tiacci, E., Falini, B. & Rabadan, R. SAVI: a statistical algorithm for variant frequency identification. *BMC. Systems Biol.* <http://dx.doi.org/10.1186/1752-0509-7-S2-S2> (2013).
39. Hubbell, E., Liu, W. M. & Mei, R. Robust estimators for expression analysis. *Bioinformatics* **18**, 1585–1592 (2002).
40. Barrett, T. *et al.* NCBI GEO: mining millions of expression profiles–database and tools. *Nucleic Acids Res.* **33**, D562–D566 (2005).
41. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article 3 (2004).
42. Simon, R. M. *et al.* in *Design and Analysis of DNA Microarray Investigations* (eds Simon, R. M. *et al.*) 65–94 (Springer, 2003).
43. van Es, J. H. *et al.* Notch/ γ -secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).

44. Eghbali-Fatourehchi, G. Z. *et al.* Circulating osteoblast-lineage cells in humans. *N. Engl. J. Med.* **352**, 1959–1966 (2005).
45. Rubin, M. R. *et al.* Parathyroid hormone stimulates circulating osteogenic cells in hypoparathyroidism. *J. Clin. Endocrinol. Metab.* **96**, 176–186 (2011).
46. Manavalan, J. S. *et al.* Circulating osteogenic precursor cells in type 2 diabetes mellitus. *J. Clin. Endocrinol. Metab.* **97**, 3240–3250 (2012).
47. Ghosh-Choudhury, N., Harris, M. A., Feng, J. Q., Mundy, G. R. & Harris, S. E. Expression of the *BMP 2* gene during bone cell differentiation. *Crit. Rev. Eukaryot. Gene Expr.* **4**, 345–355 (1994).
48. Rached, M. T. *et al.* FoxO1 is a positive regulator of bone formation by favoring protein synthesis and resistance to oxidative stress in osteoblasts. *Cell Metab.* **11**, 147–160 (2010).
49. Parfitt, A. M. *et al.* Bone histomorphometry: standardization of nomenclature, symbols, and units. Report of the ASBMR Histomorphometry Nomenclature Committee. *J. Bone Miner. Res.* **2**, 595–610 (1987).
50. Parfitt, A. M. in *Bone Histomorphometry: Techniques and Interpretation* (ed. Recker, R. R.) 143–223 (CRC Press, 1983).
51. Recker, R. R. *et al.* Issues in modern bone histomorphometry. *Bone* **49**, 955–964 (2011).



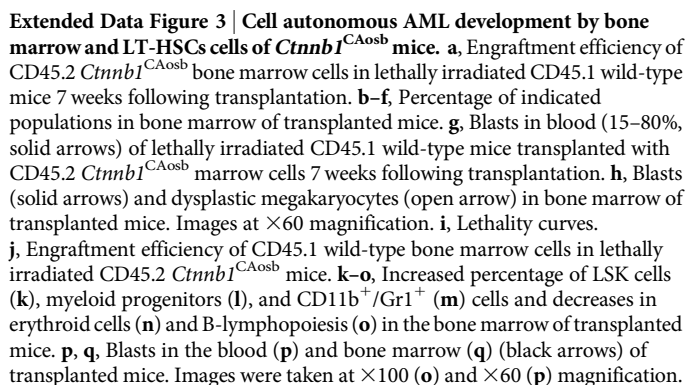
Extended Data Figure 1 | Anaemia, peripheral blood leukocytosis and monocytosis and deregulated haematopoiesis specific activation of β -catenin in osteoblasts of *Ctnnb1*^{CAosb} mice. a, Haematopoietic parameters. White blood cells (WBC), red blood cells (RBC), haemoglobin (HB), haematocrit (HCT), platelets, lymphocytes (LY), neutrophils (NE) and monocytes (MO) in 2-week-old mice. **b–k**, Bone marrow samples. **b**, Erythroid cell numbers. **c**, Representative flow cytometry image showing monocytic/granulocytic (CD11b/Gr1⁺) subset. **d**, Numbers of the CD11b⁺/Gr1⁺ subset. **e**, Distribution of LSK (Lineage[−]Sca⁺C-kit⁺) population. **f**, LSK numbers. **g**, Frequency (**g**) and percentage (**h**) of LT-HSCs and ST-HSCs. **i**, Numbers of LSK⁺/FLT3⁺ cells. **j**, Myeloid progenitor profile by CD34 versus FcγRII/III analysis of electronically gated Lin[−]Sca-1[−]c-Kit⁺ bone marrow cells. **k**, Numbers of myeloid progenitor populations. **l**, Spleen weight.

m, Extramedullary haematopoiesis in the liver of 3-week-old *Ctnnb1*^{CAosb} mice indicated by megakaryocytes (black arrow), myeloid (white arrow) and rare erythroid precursors (dotted arrow). **n–p** Percentage of Ter119⁺ (**n**), CD11b⁺/Gr1⁺ (**o**) and myeloid progenitor (**p**) populations in the spleen. **q**, PCR analysis of genomic DNA from osteoblasts and indicated haematopoietic populations from wild-type and *Ctnnb1*^{CAosb} mice. **r–t**, Real-time PCR analysis of β -catenin targets in bone marrow CD45⁺CD34⁺CD31⁺ cells (**r**), spleen (**s**) and bones (**t**). In **a**, $n = 6$; **b–k**, **n–p**, $n = 8$; **l**, $n = 5$ and **r–t**, $n = 4$ mice per group. Results are mean \pm s.d. and show a representative from five (**a–p**) or 2 (**q–t**) independent experiments. * $P < 0.05$ versus wild type. MNC, mononuclear cells.

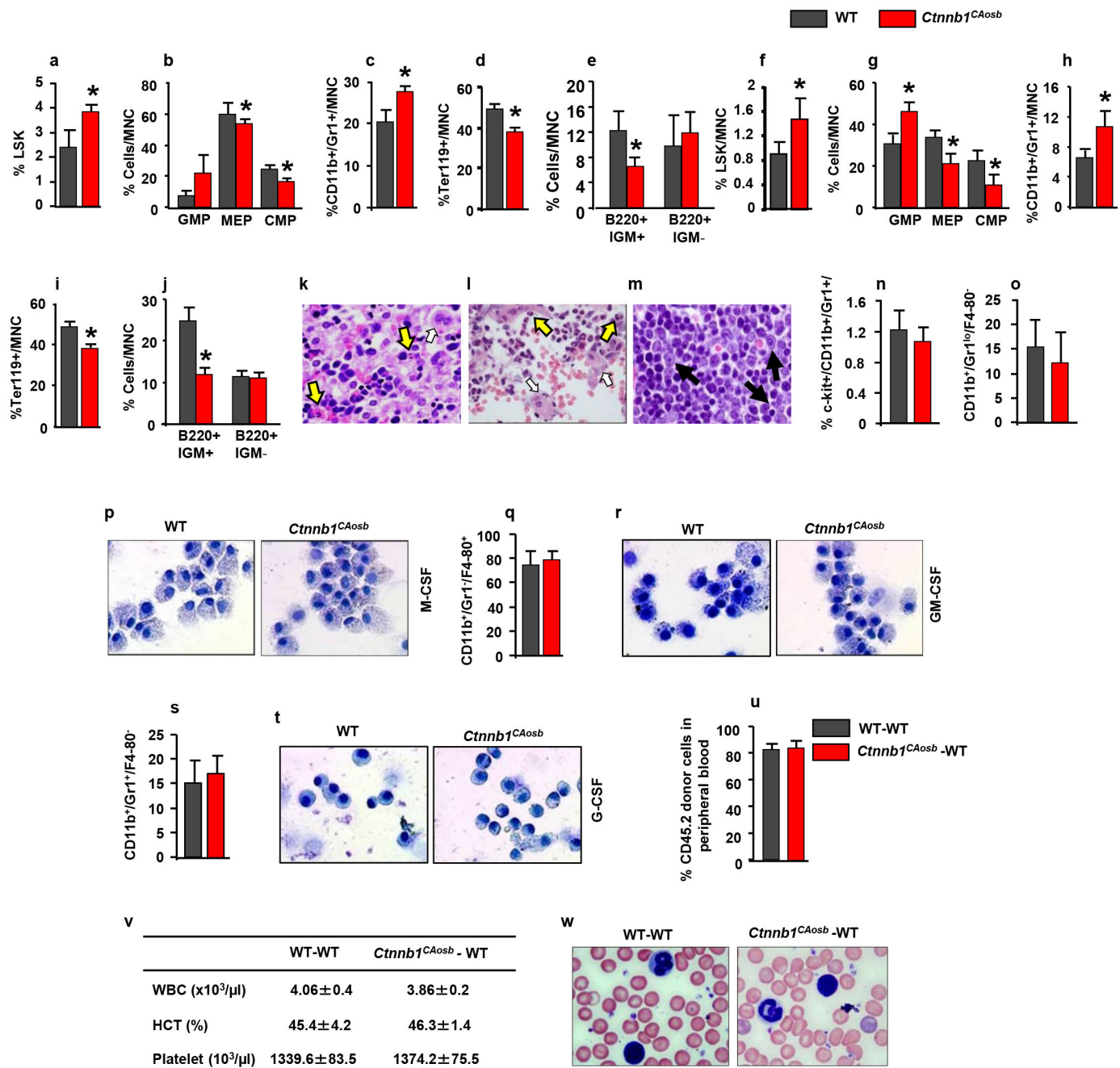


Extended Data Figure 2 | Multi-organ infiltration with blasts and dysplastic cells and myeloid differentiation block in *Ctnnb1*^{CAosb} mice. **a**, Blast infiltration (solid arrows) and neutrophil hypersegmentation (open arrow and magnified panels) in the blood of *Ctnnb1*^{CAosb} mice. Images at $\times 40$ or $\times 100$ magnification. **b**, Blast infiltration (solid arrows) and micro-megakaryocytes with hyperchromatic nuclei (white arrows) in the bone marrow of *Ctnnb1*^{CAosb} mice. Images at $\times 60$ magnification. **c**, Blast infiltration (solid arrows and magnified panel) and presence of dysplastic megakaryocytes (yellow arrows and magnified panel) in the spleen of *Ctnnb1*^{CAosb} mice. Image at $\times 400$ magnification. **d–f**, Myeloperoxidase (MPO) staining of long bone (**d**), spleen (**e**), and liver (**f**) showing massive invasion of myeloid cells. **g**, CD117 (c-kit) staining of bone sections showing CD117⁺ blasts in *Ctnnb1*^{CAosb} mice. **h**, CD13 staining of bone sections showing myeloid/monocytic

infiltration in *Ctnnb1*^{CAosb} mice. In **d–h** images at $\times 60$ magnification. **i–l**, B-cell progenitors numbers in the bone marrow (**i, j**), spleen (**k**) and lymph nodes (**l**). **m–t**, Proportion of T cells. **u**, Lack of myeloid cell differentiation in *Ctnnb1*^{CAosb} bone marrow cells following treatment with cytokines. **v–x**, Percentage of immature myeloid cells in *ex vivo* bone marrow cultures treated with cytokines. **y**, Robertsonian translocation between chromosomes 1 and 19 in 2 of 30 metaphases of the spleen of an 18-day-old *Ctnnb1*^{CAosb} mouse. Inset shows the same abnormality in another cell. **z**, Whole-exome sequencing of myeloid malignancies (CD11b⁺/Gr1⁺) from 3 *Ctnnb1*^{CAosb} mice and 3 germline normal (tail) samples. In **i–x**, $n = 6$ mice per group. $*P < 0.05$ versus wild type. Results are mean \pm s.d. and show a representative of five (**i–t**) or three (**u–x**) independent experiments.

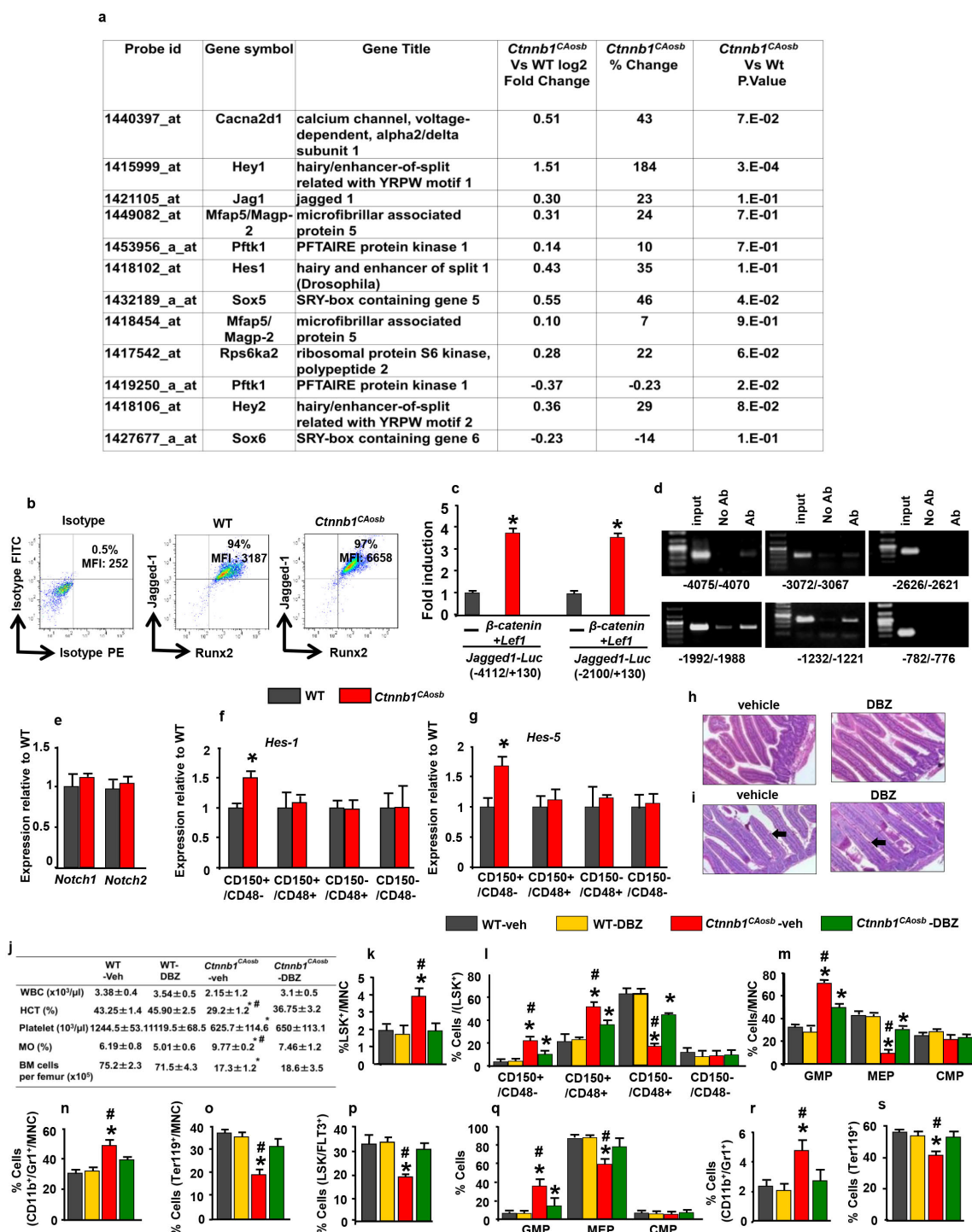


©2014 Macmillan Publishers Limited. All rights reserved



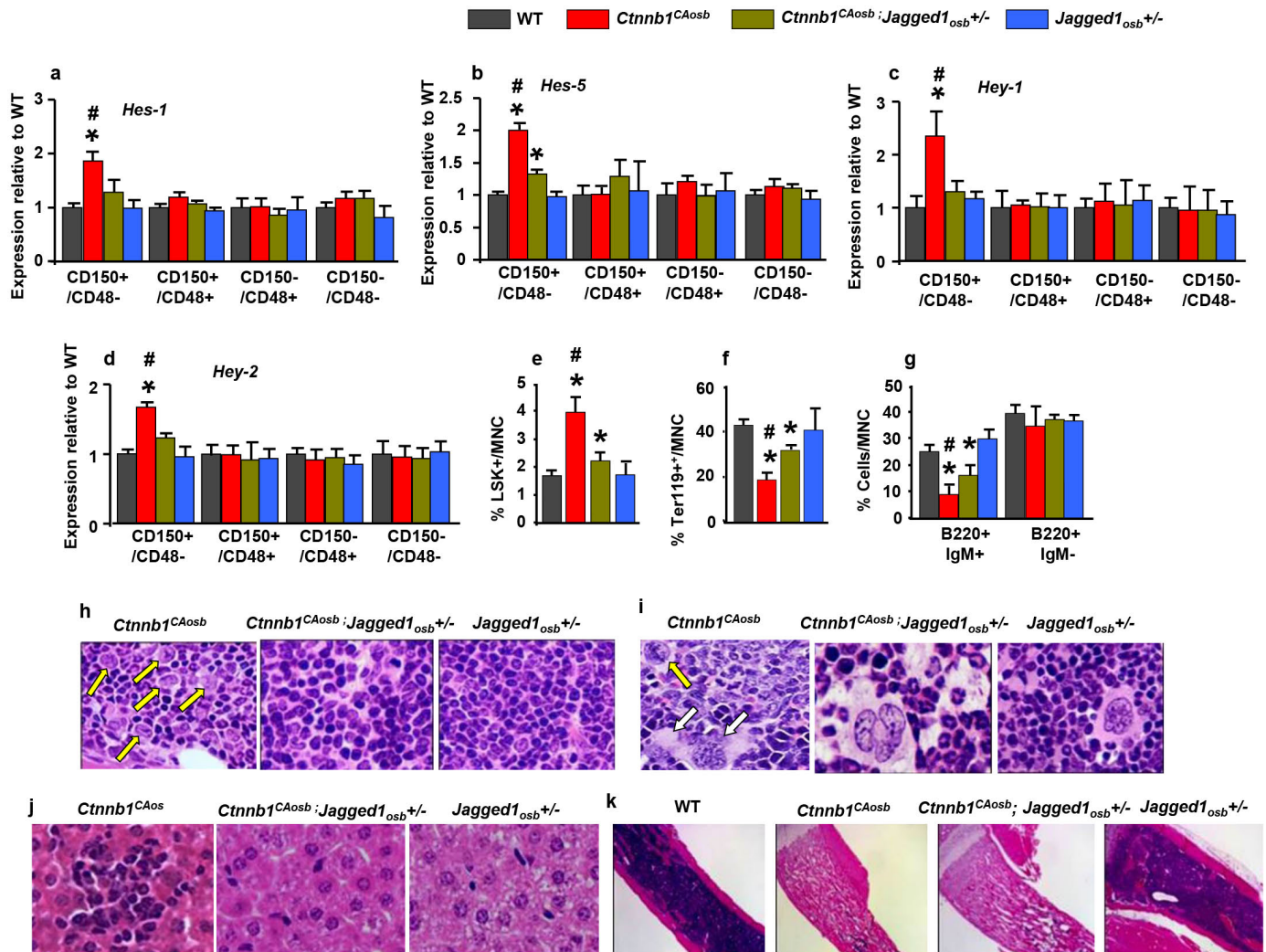
Extended Data Figure 4 | Newborn *Ctnnb1*^{CAosb} mice show MDS but fetal HSCs from *Ctnnb1*^{CAosb} mice do not transfer AML. **a–e**, Increased percentage of LSK cells (**a**), GMPs (**b**), and CD11b⁺/Gr1⁺ (**c**) cells and decreases in erythroid (**d**) cells and B-lymphopoiesis (**e**) in the liver of newborn (P1) *Ctnnb1*^{CAosb} mice. **f–j**, Increased percentage of LSK cells (**f**), GMPs (**g**), and CD11b⁺/Gr1⁺ (**h**) cells and decreases in erythroid cells (**i**) and B-lymphopoiesis (**j**) in the bone marrow of newborn (P1) *Ctnnb1*^{CAosb} mice. **k–m**, Liver (**k**), bone marrow (**l**) and spleen (**m**) of newborn (P1) *Ctnnb1*^{CAosb} mice showing microhypolobated megakaryocytes (open arrows), Pelger Huet neutrophils (yellow arrows) or blasts (solid arrows). Images at

×100 magnification. **n**, Percentage of immature myeloid cells in the bone marrow of newborn mice. **o–t**, Flow cytometry and Giemsa-stained cytopspins showing lack of changes in the percentage of immature myeloid cells in *ex vivo* cultures of bone marrow cells from P1 stage *Ctnnb1*^{CAosb} mice and treated with indicated cytokines. **u**, Engraftment efficiency of CD45.2 *Ctnnb1*^{CAosb} LSK cells obtained from the liver of embryonic day 18.5 embryos in sublethally irradiated CD45.1 wild-type mice. **v**, Normal peripheral blood measurements in transplanted mice. **w**, Lack of blasts in the blood of transplanted wild-type mice. Images at ×100 magnification. *n* = 6 mice per group. Results are mean ± s.d. and represent at least two independent experiments. **P* < 0.05 versus wild type.



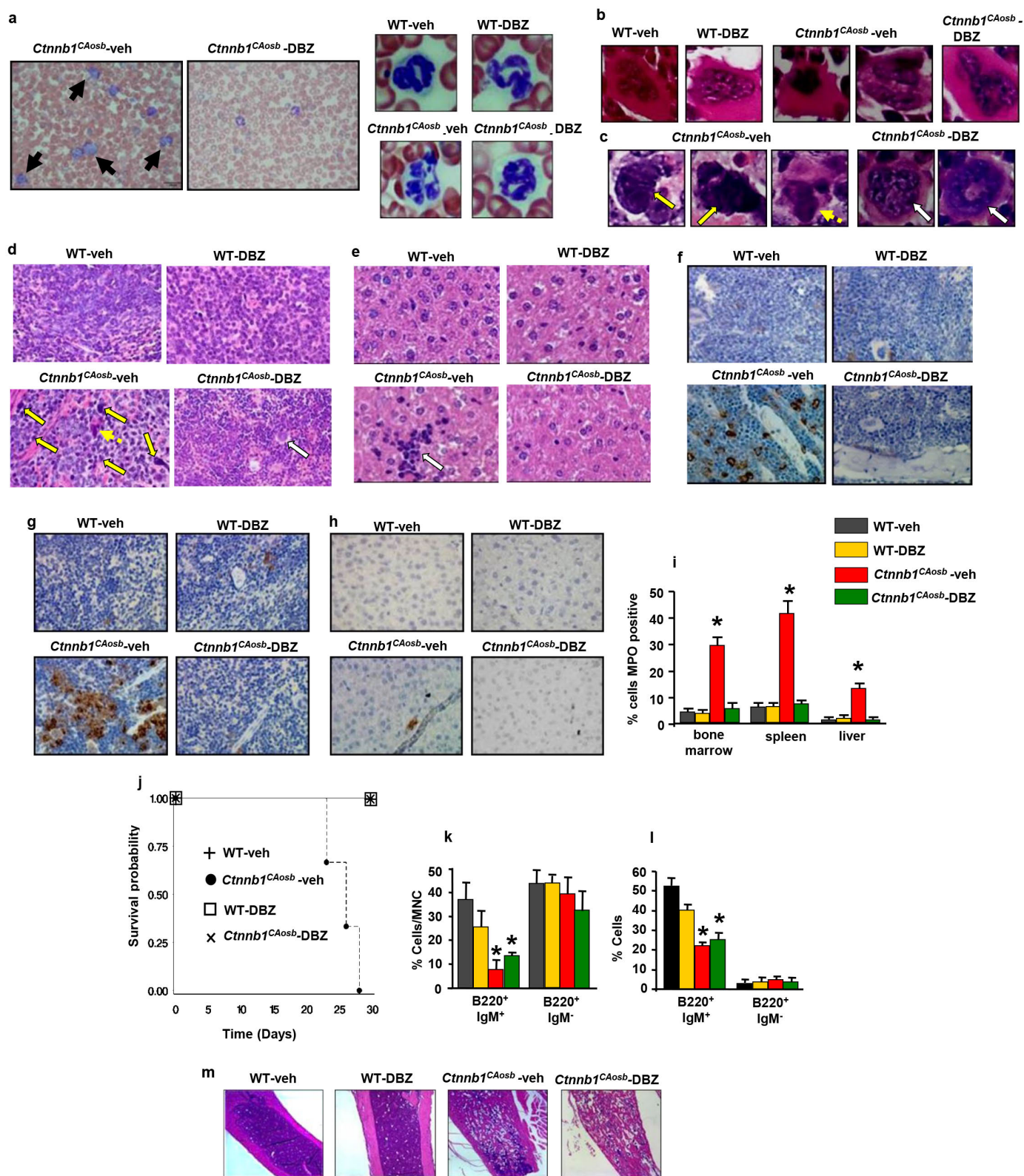
Extended Data Figure 5 | Inhibition of increased Notch signalling normalizes blood counts and rescues haematopoietic defects in *Ctnnb1*^{CAosb} mice. **a**, Microarray analysis of calvaria-derived osteoblasts from *Ctnnb1*^{CAosb} mice. AML and Notch-related genes in *Ctnnb1*^{CAosb} osteoblasts and with *P* < 0.05 and fold change of \pm 20% in one comparison. Genes that are up- or downregulated relative to wild type are shown. **b**, Flow cytometry analysis of jagged 1 expression in osteoblasts (MFI: mean fluorescent intensity). **c**, Luciferase activity in HEK293T cells co-transfected with β -catenin, *Left1* and *Jag1-Luc* reporter constructs (−4112/+130) and (−2100/+130). Results show fold induction over respective *Jag1-Luc* reporter constructs. **P* < 0.05 versus respective *Jag1-Luc*. Results are mean \pm s.d. **d**, ChIP in primary osteoblasts using anti- β -catenin antibody. Primers spanned the putative TCF/LEF binding sites (indicated) on the *Jag1* promoter. **e**, Expression of *Notch1* and *Notch2* in LSK⁺ cells. **f–g**, Expression of Notch targets in LSK⁺ subpopulations. **h, i**, Normal intestinal architecture (**h**) and PAS staining (**i**) showing lack of

goblet cell (arrows) metaplasia in DBZ-treated mice. Images at \times 60 magnification. **j**, Peripheral blood counts and bone marrow cellularity in wild-type and *Ctnnb1*^{CAosb} mice treated daily with vehicle or DBZ (2 μ mol per kg body weight) for 10 days. **k–p**, Percentage of LSK cells (**k**), LSK⁺ subpopulations (**l**), myeloid progenitors (**m**), CD11b⁺/Gr1⁺ population (**n**), erythroid cells (**o**) and LSK⁺/FLT3⁺ population (**p**) in the bone marrow. **q–s**, Percentage of myeloid progenitor populations (**q**), CD11b⁺/Gr1⁺ cells and erythroid (**s**), cells in the spleen. In **a** *n* = 3 mice per group and in **b**, *n* = 4 mice per group. In **c**, *d*, results represent two independent experiments. In **e–g** *N* = 4 mice per group, and **P* < 0.05 versus wild type. In **h–s** *n* = 8 mice per group and **P* < 0.05 versus wild type and #*P* < 0.05 *Ctnnb1*^{CAosb} vehicle versus DBZ-treated *Ctnnb1*^{CAosb} group. Results are mean \pm s.d. and show a representative of two independent experiments.



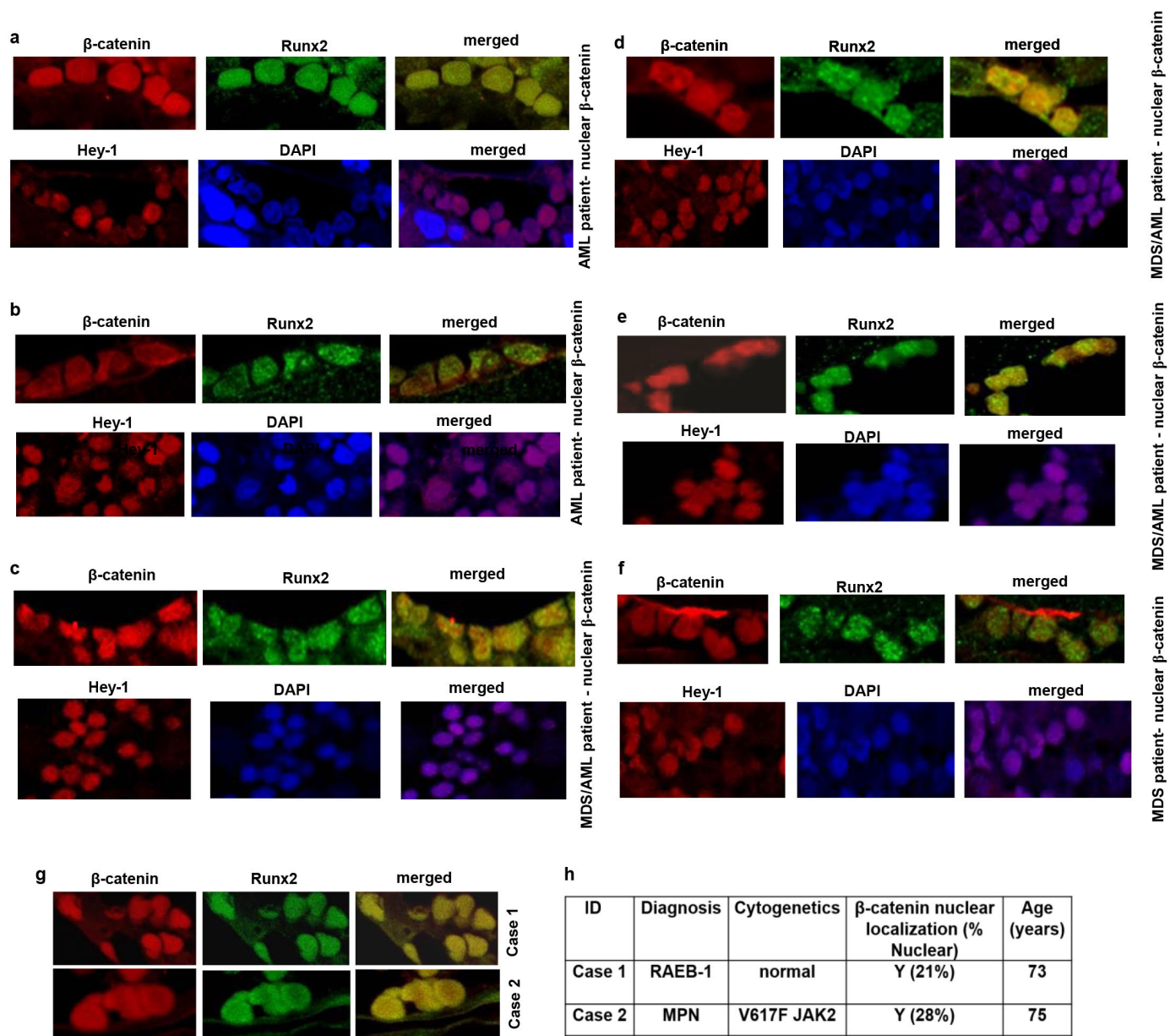
Extended Data Figure 6 | *Jag1* inactivation in osteoblasts prevents AML of *Ctnnb1*^{CAosb} mice. a–d, Expression of Notch transcriptional targets in bone marrow LSK subpopulations. Rescue of changes in the proportions of LSK (e) and erythroid (f) cells in the bone marrow of *Ctnnb1*^{CAosb}/*Jag1*^{osb}+/- mice. g, Improvement of B-lymphopoiesis in *Ctnnb1*^{CAosb}/*Jag1*^{osb}+/- mice. Normal (h) bone marrow, spleen (i) and liver (j) histology in

Ctnnb1^{CAosb}/*Jag1*^{osb}+/- mice. k, Long bone sections. Images at ×4 magnification. In a–d, *n* = 4 and in e–k *n* = 8 mice per group. **P* < 0.05 versus wild type and #*P* < 0.05 versus *Ctnnb1*^{CAosb}/*Jag1*^{osb}+/- mice. Results are mean ± s.d. and show a representative of three (a–d) and two (e–k) independent experiments.



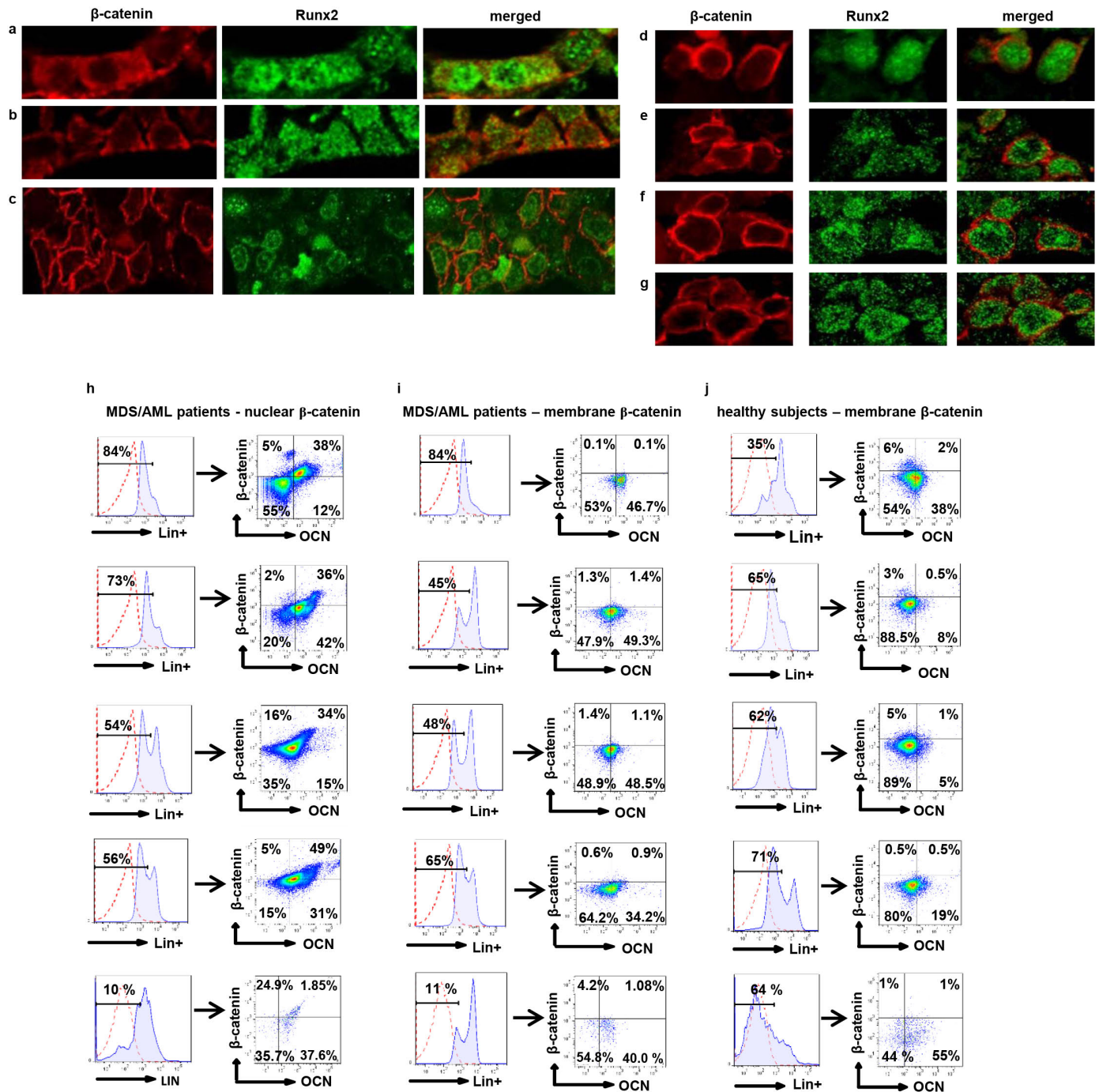
Extended Data Figure 7 | Inhibition of Notch signalling reverses AML in *Ctnnb1*^{CAosb} mice. **a**, Lack of blasts (solid arrows) and, normal neutrophils (right panel) in blood of DBZ-treated *Ctnnb1*^{CAosb} mice. **b–c**, Normal megakaryocytes in the bone marrow (**b**) and spleen (**c**). **d**, Normal spleen histology in DBZ-treated *Ctnnb1*^{CAosb} mice. Yellow arrows indicate abnormal cells with large nucleoli and dotted arrow indicates abnormal megakaryocytes in *Ctnnb1*^{CAosb} mice; white arrow indicates normal megakaryocytes in DBZ-treated *Ctnnb1*^{CAosb} mice. **e**, Lack of monocyte infiltration in the liver of DBZ-treated *Ctnnb1*^{CAosb} mice. Arrow indicates cluster of mononuclear cells.

f–h, MPO staining of bone marrow (**f**), spleen (**g**) and liver (**h**). **i**, Per cent of cells staining with MPO in the indicated tissues. **j**, Increased survival in DBZ-treated *Ctnnb1*^{CAosb} mice. In **a**, **b**, images taken at $\times 100$ magnification. In **c–g** images taken at $\times 60$ magnification. **k–l**, Proportion of B-cell populations in the bone marrow (**k**) and spleen (**l**). **m**, Long bone sections. Images at $\times 4$ magnification. $n = 6$ mice per group. $*P < 0.05$ versus wild type and $\#P < 0.05$ *Ctnnb1*^{CAosb} vehicle versus DBZ-treated *Ctnnb1*^{CAosb} group. Results are mean \pm s.d. and show a representative from two independent experiments.



Extended Data Figure 8 | Nuclear accumulation of β-catenin in osteoblasts and increased Notch signalling in 38.3% of patients with MDS/AML and identification of underlying pre-AML conditions by nuclear localization of β-catenin in osteoblasts. **a–f**, Double immunofluorescence staining with β-catenin and RUNX2 in osteoblasts from bone marrow biopsies from 6 MDS/AML patients harbouring nuclear accumulation of β-catenin in osteoblasts and showing nuclear accumulation of HEY1 in the corresponding patients (×60 magnification). **g**, **h**, During screening assumed healthy controls, 2 individuals

were identified with nuclear β-catenin in their osteoblasts. Re-evaluation showed underlying hematologic disorder, Case 1: MDS RAEB-1, Case 2: Jak2 positive myelofibrosis. **g**, Double immunofluorescence staining with β-catenin and RUNX2 in osteoblasts from bone marrow biopsies of the 2 cases (×60 magnification). **h**, β-catenin cellular localization in cases 1 and 2 with associated cytogenetic abnormalities. NL: normal cytogenetics. In the fourth column percentages indicate osteoblasts with nuclear localization of β-catenin.



Extended Data Figure 9 | Membrane accumulation of β -catenin in osteoblasts in 61.7% of patients with MDS/AML and in healthy subjects and nuclear accumulation of β -catenin in osteoblasts in 38.3% of patients with MDS/AML identified by flow cytometry. Double immunofluorescence staining with β -catenin and Runx2 in osteoblasts from bone marrow biopsies from **a–c**, 3 MDS/AML patients and **d–g**, 4 healthy subjects harbouring membrane localization of β -catenin in osteoblasts. **h–j**, Flow cytometry using a

non-phospho β -catenin antibody detecting nuclear/activated β -catenin. Representative plots showing nuclear (**h**), versus non-nuclear (**i**), localization of β -catenin in osteoblasts from individual MDS/AML patients and, non-nuclear localization of β -catenin in osteoblasts from 5 healthy subjects as $CD34^- / Lin^- OCN^+$ cells, (OCN, osteocalcin an osteoblast-specific protein used for isolation of live osteoblastic cells).

Extended Data Table 1 | Structural chromosomal alterations detected by aCGH in *Ctnnb1*^{CAosb} mice

Mouse Chromosome	Location (Mb)	Size (kb)	Genes	Copy number change	Mouse ID	Human chromosome	Location (Mb)
2	54.7	114	Glnt13	Deletion of 2/3 copies	1	No ortholog	
3	93.4	608	Tdpz2, Tdpz1, Tdpz4, Tdpz3, Tdpz5, Gm10697, Gm9125, Gm10696	Deletion of 1/3 copies	1	No orthologs	
4	62.1	245	Bspry, Hdh3, Alad	Deletion of 1/3 copies	3	9q32	115.16 115.17 115.2
5	26.3	980	Speer4a, BC061212, A430089119Rik	Deletion of 3/4 copies	2, 3	No orthologs	
	95.6	238	AA792892, Gm16367, D5Etd577e, 100041354	Deletion of 1/3 copies	5		
5	5.1	196	Pfkf1	Gain of < 1 copy	2	7q21.13	90.4
	13.5	181	Sema3a	Gain of 1 copy	2, 3	7q21.11	83.5
	15.7	398	Cacna2d1	Gain of < 1 copy	2, 3	7q21.11	81.6
	22.6	227	Lhpl3	Gain of > 1 copy	2, 3, 4, 5	7q22.1	104
6	144	59	Sox5	Deletion of 3/4 copies	1, 2, 3, 4	12p12.1	23.9
7	123	59	Sox6	Deletion of 2/3 copies	5	11p15.2-15.1	16.2
8	87.2	4.1	Nfix	Deletion of 1/2 copies	3, 4	19p13.2	13.2
	97.5	43	Gpr56	Deletion of 1/2 copies	3	No ortholog	
8	75	16	Slc35e1	Gain of <1 copies	3	19p13.11 19p13.11	16.5 16.6
	22.3	494	Med26	Gain of <1 copies	4	No ortholog	6.8 6.9
			Defa-rs7			No ortholog	
			Defa23			No ortholog	
9	53.3	0.06	Defa25	Heterozygous deletion	1	No ortholog	107.5
			Defa22			No ortholog	
			Defa3			No ortholog	
			Defa5			No ortholog	
10	106.2	0.06	Defcr-rs1	Heterozygous deletion	1	No ortholog	80.4
			Defa20			No ortholog	
11	51.9	12.9	Npat	Heterozygous deletion	1	11q22.3	107.5
	66.6	32	Ppfa2	Heterozygous deletion	1, 3, 4	12q21.31	80.4
	90.1	5.9					
12	8.9	0.05	Matn3	Heterozygous deletion	5	5q31.1	133.5
	114.7	475	Adam6b, Adam6a	Deletion of 2/3 copies	4, 5		
			Sntg2				
12	31	71	No genes	Deletion of 1/2 copies	4	14q32.33	105.5
13	67.5	163	No genes	Deletion of 2/5 copies	1	2p25.3	1.1
14	108	380	No genes	Heterozygous deletion	3		
	110	19	1700112E06Rik	Deletion of 2/3 copies	1, 2, 3, 4		
15	23.7	201	No genes	Heterozygous deletion	5	No ortholog	
15	20.4	20	No genes	Deletion of 2/3 copies	2, 3, 4		
	77.3	47	Apol10a, Apol7c	Gain of < 1 copy	2, 3, 4, 5	No ortholog	
17	15.1	86	Gm3448, Gm3417, 9030025P20Rik, Gm3435, Tcte3, 2210404J11Rik	Heterozygous deletion	2, 3, 4, 5	No ortholog	170.1
	35.4	38	LOC547349	Deletion of 2/3 copies	2, 5	No ortholog	49.8 167.8
		4	H2-T3, H2-Tw3			No ortholog	
		60	Crisp1			No ortholog	
	36.3	494	Tcp10a, Tcte2, Mllt4	Deletion of 3/4 copies	2, 5	6p12.3	
	40.4		Zfp472, C920016K16Rik	Deletion of 1/3 copies	4	6q27	168
	33.1	33	Olfir137, Olfir136	Deletion of 2/5 copies	4	No ortholog	
	38.4	60		Deletion of 1 copies	4	No ortholog	
						No ortholog	
17	6.1	701	Tulp4, Tmem181a, Dynl1, Tmem181b, Tmem181c-ps, Tmem181d-ps, Sytl3	Gain of < 1 copy	2	6q25.3	158.7 158.9 159
	7.4	45	Rps8ka2, Tcp10b, Tcp10a	Gain of < 1 copy	2, 3	6q27 6q27	159 166.9 167.7
		65	H2-Q8, H2-Q6, H2-Q7			No ortholog	
	35.5		H2-T24, H2-T23, H2-T9, H2-T22, EG547347, H2-BI, H2-T10, Gm8909, H2-T3, H2-Tw3	Gain of < 1 copy	2	No ortholog	
	36.1	241	No genes	Gain of < 1 copy	2, 3, 5	No ortholog	
	37.5	19	Olfir137, Olfir136	Gain of >1 copy	2, 5	No ortholog	
	38.4	60		Gain of >2 copies	2, 3, 5	No ortholog	
	13.4	364	Tcp10a	Deletion of 2/5 copies	3, 5	6q27	167.7
19	37.4	114	Ide	Deletion of 2/5 copies	3, 4	10q23.33	94.3
	36.9	39	Fgfbp3, Btaf1	Deletion of 1/2 copies	3	10q23.32	93.6, 93.7

Mouse ID denotes mouse number (1, mouse 1; 2, mouse 2 and so on).

RNA viruses can hijack vertebrate microRNAs to suppress innate immunity

Derek W. Trobaugh¹, Christina L. Gardner¹, Chengqun Sun¹, Andrew D. Haddow², Eryu Wang², Elik Chapnik³, Alexander Mildner⁴, Scott C. Weaver², Kate D. Ryman¹ & William B. Klimstra¹

Currently, there is little evidence for a notable role of the vertebrate microRNA (miRNA) system in the pathogenesis of RNA viruses¹. This is primarily attributed to the ease with which these viruses mutate to disrupt recognition and growth suppression by host miRNAs^{2,3}. Here we report that the haematopoietic-cell-specific miRNA miR-142-3p potentially restricts the replication of the mosquito-borne North American eastern equine encephalitis virus in myeloid-lineage cells by binding to sites in the 3' non-translated region of its RNA genome. However, by limiting myeloid cell tropism and consequent innate immunity induction, this restriction directly promotes neurologic disease manifestations characteristic of eastern equine encephalitis virus infection in humans. Furthermore, the region containing the miR-142-3p binding sites is essential for efficient virus infection of mosquito vectors. We propose that RNA viruses can adapt to use antiviral properties of vertebrate miRNAs to limit replication in particular cell types and that this restriction can lead to exacerbation of disease severity.

miRNAs are 21–23-nucleotide host-encoded RNAs that are cell-specific and bind to complementary sequences in the 3' non-translated region (NTR) of host messenger RNAs (mRNAs)⁴. The extent of sequence complementarity between the miRNA and mRNA leads to control of mRNA-encoded polypeptide levels by either a block in translation, degradation of the mRNA, or both^{5,6}. For RNA viruses, limited evidence exists for host miRNAs binding to viral RNAs and restricting infection or affecting disease^{1,7,8}. In the case of hepatitis C virus, the opposite is observed: the liver-specific miRNA miR-122 binds to the viral 5' NTR, stabilizing the RNA and enhancing *in vitro* viral replication^{9,10}.

Wild-type North American eastern equine encephalitis virus (EEEV) strains are highly virulent mosquito-borne alphaviruses causing a 30–70% case fatality rate in humans¹¹. The recognized geographic range and disease incidence of EEEV in the northeastern United States has increased over the past 10 years, raising concern about potential widespread outbreaks¹². EEEV disease is characterized by a limited prodrome before manifestations of encephalitis, resulting from restricted myeloid cell replication and minimal induction of systemic type I interferon (IFN)^{13,14}. Longer prodromes in human paediatric cases increase the likelihood of recovery, suggesting that host prodromal responses may limit disease severity¹⁵.

Wild-type EEEV is defective for replication in human and murine macrophages and dendritic cells¹³. Using a luciferase-expressing translation reporter RNA encoding the 5' and 3' NTRs and translation initiation control sequences of wild-type EEEV (Extended Data Fig. 1a), we found that translation was restricted in murine RAW 264.7 (RAW) cells, a monocyte/macrophage myeloid cell line, versus BHK-21 fibroblasts (Fig. 1a and Extended Data Fig. 1d)¹³. Translation of an analogous reporter RNA derived from the related myeloid-cell-tropic wild-type Venezuelan equine encephalitis virus (VEEV) was efficient in both RAW (Fig. 1a) and BHK-21 (Extended Data Fig. 2a, b) cells^{13,16}. Removal of the EEEV 5' NTR (EEEV 5'Δ NTR) (Extended Data Fig. 1b) did not

alleviate the restriction in translation in RAW cells (Fig. 1a), suggesting that the EEEV 3' NTR confers this restriction. Indeed, transfer of the EEEV 3' NTR to a host mRNA mimic (5' host 3' EEEV) (Extended Data Fig. 1c) resulted in translation blockade in RAW cells but not in BHK-21 cells (Fig. 1a and Extended Data Fig. 1d). Transfer of the VEEV 3' NTR to the host mimic had no effect on translation in RAW or BHK-21 cells (Extended Data Fig. 2a, b). Therefore, the EEEV 3' NTR but not the VEEV 3' NTR contains the restricting element(s).

Two miRNA prediction algorithms, miRANDA¹⁷ and PITA¹⁸, identified three putative canonical and one non-canonical binding sites for the haematopoietic-cell-specific miRNA miR-142-3p in the 3' NTR of the North American EEEV strain FL93-939 (Extended Data Fig. 3a, b). The three canonical miR-142-3p seed sites are conserved in 17 out of 23 sequenced North American EEEV strains collected between 1954 and 2012, suggesting a strong selection for their retention¹⁹ (S.C.W., unpublished observations). To determine whether the miR-142-3p binding sites in the EEEV 3' NTR restrict viral replication, we generated an EEEV mutant (11337) with a deletion of 260 nucleotides encompassing all of the miR-142-3p binding sites (Extended Data Fig. 3c). In BHK-21 cells, we observed no significant difference in viral replication at 12 h post-infection (h.p.i.) with 11337 compared to wild-type EEEV ($P > 0.2$, Extended Data Fig. 3d). However, replication of 11337 in RAW cells (Fig. 1b) was nearly 1,000-fold higher than wild-type EEEV within 8 h.p.i. A similar phenotype was observed after infection of human K562 and THP-1 monocyte/macrophage cells with wild-type EEEV and 11337 (Extended Data Fig. 4a, b) as well as infection of primary bone-marrow-derived dendritic cells (BMDCs), reported to express high levels of miR-142-3p²⁰ (Fig. 1c). Wild-type EEEV remained replication-defective in BMDCs in the absence of type I IFN signalling (Extended Data Fig. 4c)¹³; however, replication of 11337 increased, indicating that myeloid cell restriction of wild-type EEEV is IFN-independent but dependent upon the 3' NTR sequences containing the miR-142-3p binding sites. All monocyte/macrophage cells used in this study expressed high levels of miR-142-3p in contrast to BHK-21 cells, in which miR-142-3p expression was undetectable (Extended Data Fig. 5).

To confirm a specific role for miR-142-3p in restricting EEEV replication, we expressed miR-142 in BHK-21 cells and assessed its effects on infection by wild-type EEEV, 11337 and wild-type VEEV. Ectopic expression of miR-142 in BHK-21 cells (Extended Data Fig. 5) completely blocked wild-type EEEV infection in comparison to control cells expressing a neuron-specific miRNA, miR-124 (Fig. 1d)²¹. By contrast, both 11337 and wild-type VEEV infected miR-142-expressing BHK-21 cells. To demonstrate the dependence of this restriction on the specific miR-142-3p binding sites, we generated mutant viruses that had either each of the miR-142-3p binding sequences deleted (142del) or three point mutations in each miR-142-3p binding site seed sequence (142pm; Extended Data Fig. 6a, b). These viruses replicated equally well in BHK-21 (Extended Data Fig. 6c) and RAW (Fig. 1e) cells, similar to 11337. To confirm that the increase in replication in myeloid cells was due to

¹Center for Vaccine Research and Department of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. ²Institute for Human Infections and Immunity, Center for Biodefense and Emerging Infectious Diseases, and Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555, USA. ³Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel. ⁴Department of Immunology, Weizmann Institute of Science, Rehovot 7610001, Israel.

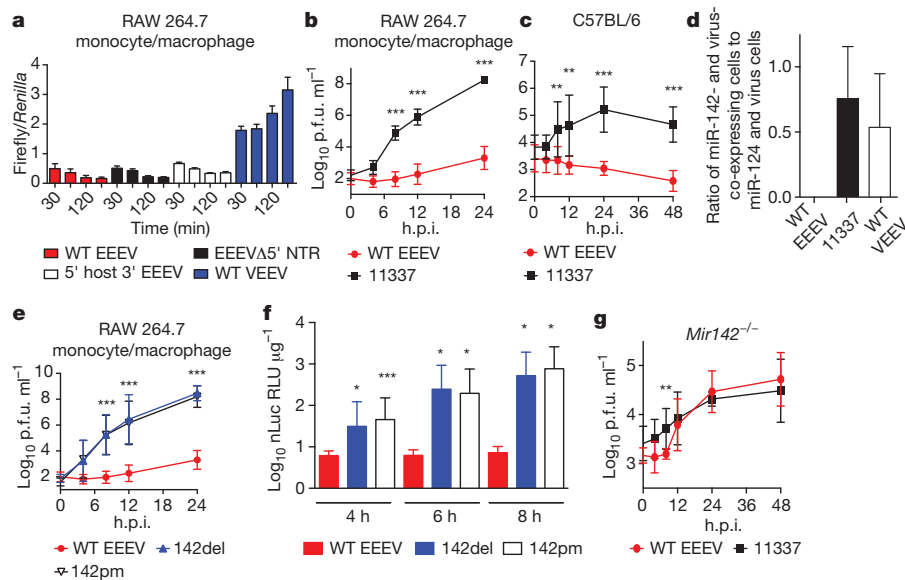


Figure 1 | EEEV restriction in myeloid cells is due to miR-142-3p binding sites in the 3' NTR. **a**, The EEEV 3' NTR restricts translation in RAW cells. Errors bars represent mean \pm standard deviation (s.d.) from three independent experiments. WT, wild type. **b**, **c**, Replication of wild-type EEEV and 11337 in RAW cells (**b**) and C57BL/6 BMDCs (**c**). p.f.u., plaque-forming units. **d**, Overexpression of miR-142-3p in BHK-21 cells blocks EEEV infection compared to overexpression of the control miR-124. Data are represented as the ratio of the percentage of cells co-expressing the microRNA miR-142 (eGFP) and virus-infected (mCherry) to cells co-expressing miR-124 and

virus-infected. Data are averaged (mean \pm s.d.) from two independent experiments. **e**, Replication of EEEV 142del and 142pm in RAW cells. **f**, Ablation of the miR-142-3p binding sites in EEEV increases translation in RAW cells. nLuc, NanoLuc-luciferase; RLU, relative light units. **g**, Replication of wild-type EEEV and 11337 are similar in *Mir142*^{-/-} BMDCs. Data represent the geometric mean \pm s.d. from three independent experiments unless indicated. Asterisks indicate differences that are statistically significant (* P < 0.05, ** P < 0.01, *** P < 0.001).

translation of the viral genome, we infected RAW cells with nsP3 reporter viruses to measure translation of virus-particle-delivered genomes¹³. Translation was significantly increased by 4 h.p.i. with 142del and 142pm viruses compared to wild-type EEEV (Fig. 1f). Finally, we infected BMDCs derived from miR-142-deficient (*Mir142*^{-/-}) mice²⁰ with wild-type EEEV or 11337, and detected no significant difference in viral titres between the viruses after 12 h.p.i. (P > 0.1, Fig. 1g). These data demonstrate that the presence of haematopoietic-cell-specific miR-142-3p binding sites in the wild-type EEEV 3' NTR results in potent blockade of viral translation and subsequent replication in miR-142-3p-expressing cells *in vitro*.

Wild-type-EEEV-infected mice exhibit a minimal prodrome (for example, ruffled fur, hunched posture, weight loss), which is probably due to restricted myeloid cell replication and minimal type I IFN induction^{13,14}. To assess the contribution of the miR-142-3p binding sites in EEEV to this phenotype, we infected CD-1 mice with wild-type EEEV or 11337. Survival times were extended in 11337-infected mice compared to wild-type-infected mice (Fig. 2a) with evidence of prodromal disease developing only in 11337-infected mice (Fig. 2b). Wild-type EEEV naturally binds heparan sulphate, which limits viral dissemination while also increasing neurovirulence¹⁴. Therefore, we included a heparan-sulphate-binding-defective EEEV mutant (71-77) for a comparison with wild-type VEEV, which does not bind heparan sulphate efficiently^{14,22}. The 71-77 mutant was significantly attenuated compared to wild-type EEEV, and elicited signs of prodrome similar to 11337, suggesting that both heparan sulphate binding and miRNA restriction contribute to the inhibition of prodromal disease. Combining the heparan-sulphate-binding-defective mutation with the 11337 deletion (71-77/11337) increased both survival times and prodromal signs compared to wild-type EEEV or 11337 infection. The timing of prodrome onset following 71-77/11337 infection was only slightly delayed compared to wild-type-VEEV-infected mice (Fig. 2b). Consistent with limited prodrome, systemic type I IFN induction is rarely detected during wild-type EEEV infection (Fig. 2c)¹³. However, IFN- α/β was detected by 12 h.p.i. in sera from 11337-infected mice, similar to mice infected with

the heparan sulphate binding mutant 71-77, and both were significantly higher than wild-type EEEV. Infection with 71-77/11337 elicited higher levels of IFN- α/β within 8 h.p.i. compared to wild-type EEEV, and at 12 h.p.i. induced IFN- α/β levels similar to those in sera from wild-type-VEEV-infected mice. These results are consistent with a study in which the addition of artificial miR-142-3p binding sites into the genome of influenza virus reduced IFN- α/β induction *in vivo*, suggesting that myeloid cell replication may be necessary for serum IFN- α/β induction with multiple viruses²³.

Previously, we observed that serum levels of IFN- α/β and prodromal signs were associated with alphavirus replication in myeloid cells within popliteal lymph nodes (PLN) after footpad inoculation^{13,14}. Although wild-type EEEV replicated poorly in the PLN compared to wild-type VEEV throughout infection (Fig. 2d), both 71-77 and 11337 viruses replicated significantly more than wild-type EEEV by 12 h.p.i.. The double mutant, 71-77/11337, replicated more efficiently in PLNs at all time points compared to either 71-77 or 11337, and at levels comparable to, but lower than, those seen in wild-type-VEEV-infected PLNs. Replication restriction was alleviated for 71-77 and 71-77/11337 in type I IFN receptor-deficient *Ifnar1*^{-/-} mice (Fig. 2d), and survival times for the four viruses were essentially identical (P > 0.1; Extended Data Fig. 7), indicating that IFN- α/β is the primary attenuating factor for 11337 and 71-77/11337, and that the 11337 mutation does not compromise replication *in vivo* in comparison with the wild-type virus. Fluorescence microscopy and flow cytometric analysis (FCA) of PLNs from mice infected with mCherry-expressing viruses demonstrated that the number of cells infected after wild-type EEEV infection were not significantly different from uninfected mice, but 11337 and 71-77/11337 infected a significantly higher number of cells compared to wild-type EEEV, with the number of cells infected with 71-77/11337 approaching those seen after wild-type VEEV infection (Fig. 2e, f). Cells infected by 11337, 71-77/11337 and wild-type VEEV were predominantly CD11b⁺ with a subset also CD11c⁺, indicative of myeloid lineage cells (Extended Data Fig. 8). 71-77 appeared to infect some PLN cells by microscopy (Fig. 2e); however, failure to detect 71-77 PLN infection by

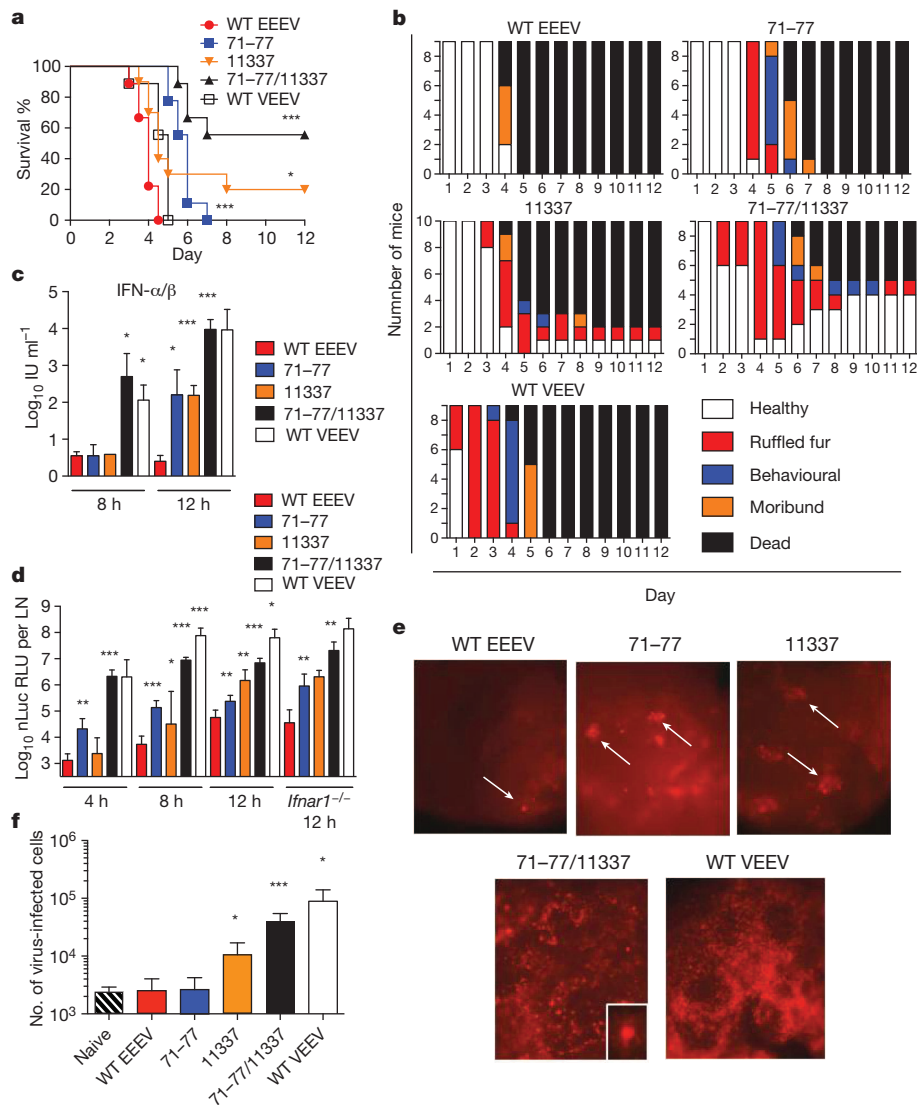


Figure 2 | miR-142-3p binding sites in EEEV 3' NTR decrease virus replication in the lymph node and enhance disease progression.

a, Survival curves of CD-1 mice. $n = 9$ and 10 (11337) mice per group, three independent experiments. **b**, Mice were monitored daily for clinical signs of disease. Y-axis represents number of mice exhibiting each sign of disease on each day. **c**, Serum levels of IFN- α/β in infected CD-1 mice. $n = 8$ mice, two independent experiments. Limit of detection of IFN- α/β ranged from 1.9 international units (IU) ml^{-1} to 3.9 IU ml^{-1} . **d**, Quantification of viral replication in the PLNs using nLuc-reporter viruses. $n = 8$ mice (CD-1) and 6 mice (*Ifnar1* $^{-/-}$) per time point, two independent experiments. **e**, Visualization of mCherry-reporter-virus-infected whole PLN from CD-1 mice 12 h.p.i. Images are representative of one PLN from two mice. Arrows indicate virus-infected cell(s) in wild-type-EEEV-, 71-77- and 11337-infected PLNs. **f**, Quantification of virus-infected cells in PLN collected 12 h.p.i. $n = 4$ (naive), 5 (71-77) or 6 (71-77/11337) mice, two independent experiments. Error bars for all experiments represent geometric mean \pm s.d. Asterisks indicate differences for all experiments that are statistically significant (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

FCA (Fig. 2f) is consistent with the presence of miR-142-3p binding sites in the 3' NTR of this virus. Overall, deletion of the miR-142-3p binding sites alone or in combination with disruption of the heparan-sulphate-binding ability of EEEV increased infection of PLN myeloid-lineage cells, prodromal disease signs, and type I IFN production but decreased virulence *in vivo* dependent upon a functioning IFN- α/β response.

Given the rapid mutation rate of RNA viruses, we proposed that the miR-142-3p binding sites in the EEEV 3' NTR are maintained through positive selection during the mosquito-vertebrate transmission cycle. In C6/36 mosquito cells, the replication of 11337, 142del and 142pm viruses was significantly reduced compared to wild-type EEEV at 12 h.p.i (Fig. 3a); however, by 24 h.p.i, only the 11337 and 142del viruses remained attenuated. Furthermore, reduced infection rates of the EEEV bridge vector, *Ochlerotatus taeniorhynchus* (also known as *Aedes taeniorhynchus*)²⁴, via artificial blood meals were observed for 11337, 142del and 142pm viruses compared to wild-type EEEV (Fig. 3b). Therefore, specific sequences of the miR-142-3p binding sites are required for efficient mosquito infection.

We have demonstrated that host miRNA restriction of EEEV replication in myeloid cells is a novel mechanism that determines virus tropism for this cell lineage. Moreover, at least portions of these NTR sequences promote mosquito vector infection, suggesting positive selection as a mechanism for binding-site retention during natural EEEV transmission. It is also clear from these data that, at the organism level,

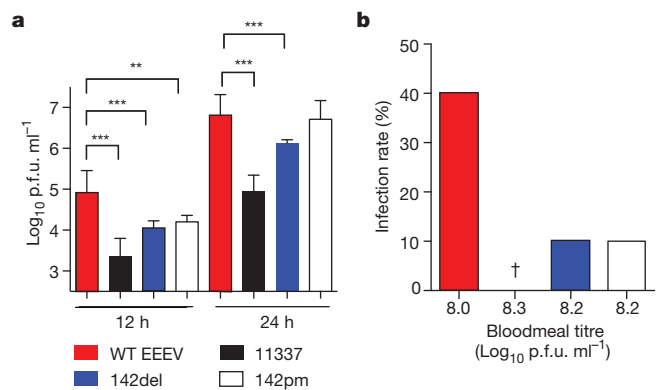


Figure 3 | EEEV sequences containing the miR-142-3p binding sites in EEEV are required for efficient mosquito infection. **a**, Replication of wild-type EEEV, 11337, 142del and 142pm in C6/36 mosquito cells. $n = 3$ or 4 (11337) independent experiments. Error bars represent geometric mean \pm s.d. and asterisks indicate differences that are statistically significant using a two-tailed unpaired *t*-test comparing wild-type EEEV to all other viruses (** $P < 0.01$, *** $P < 0.001$). **b**, Infection rates of *Aedes taeniorhynchus* after ingestion of infectious bloodmeals ($n = 20$ mosquitoes per virus). † indicates 0 of 20 mosquitoes infected.

miRNA-mediated restriction of virus replication can lead to suppression of innate immune responses and exacerbation of disease, thereby benefiting the infecting microorganism. Understanding the role of miRNA expression levels and virus genotype in the efficiency of restriction may provide insight into temporal, geographical and individual host variation in EEEV and potentially other RNA virus diseases.

METHODS SUMMARY

Culture of BHK-21, L929 fibroblasts and RAW monocyte/macrophage cells and BMDCs from CD-1, *Ifnar1*^{-/-} and *Mir142*^{-/-} mice has been described^{13,20}. Virus growth curves with BHK-21 cell titration of progeny viruses were performed as described previously¹³ using enhanced GFP (eGFP) or mCherry reporter viruses described below.

Wild-type EEEV, wild-type VEEV and host mimic luciferase-expressing translation reporters were described previously^{25,26}. Other reporters were constructed using the QuikChange II XL mutagenesis kit. Translation assays were performed as described with minor modifications¹³.

Construction of cDNA clones of VEEV ZPC 738 (ref. 27; wild-type VEEV), EEEV FL93-939 (ref. 28; wild-type EEEV) and EEEV 71-77 (ref. 14) were described previously. The EEEV 11337, 142del and 142pm mutants were constructed from the FL93-939 cDNA. mCherry-, eGFP- and nLuc-expressing versions of all viruses were constructed similarly to a described capsid-PE2 fusion reporter viruses using the QuikChange II XL kit²⁹.

The eGFP-expressing pCMV-miR-142 or pCMV-miR-124 expression plasmids were electroporated into BHK-21 cells. After ~18 h, cells were infected with mCherry reporter viruses for 8 h followed by assessment of co-expression of the miRNA (eGFP) and virus reporter (mCherry) using fluorescence microscopy.

Outbred CD-1 mice were infected subcutaneously and evaluated for morbidity and mortality as described^{13,14}. For tissues, CD-1 or *Ifnar1*^{-/-} mice were infected with nLuc or mCherry reporter viruses. PLNs were collected at indicated time points and processed for Nano-Glo luciferase assay, for fluorescence microscopy, or for FCA of virus-infected cells. Serum IFN- α/β was measured using a standard biological assay as described¹³.

Adult female *A. taeniorhynchus* mosquitoes were infected with EEEV viruses in artificial blood meals. Engorged females were incubated 10 days and assayed for infection by infection of Vero cells and observation for cytopathic effects³⁰.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 July; accepted 5 November 2013.

Published online 18 December 2013.

1. Cullen, B. R. MicroRNAs as mediators of viral evasion of the immune system. *Nature Immunol.* **14**, 205–210 (2013).
2. Pham, A. M., Langlois, R. A. & tenOever, B. R. Replication in cells of hematopoietic origin is necessary for Dengue virus dissemination. *PLoS Pathog.* **8**, e1002465 (2012).
3. Zheng, Z. *et al.* Human microRNA hsa-miR-296-5p suppresses enterovirus 71 replication by targeting the viral genome. *J. Virol.* **87**, 5645–5656 (2013).
4. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
5. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
6. Djuranovic, S., Nahvi, A. & Green, R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**, 237–240 (2012).
7. Otsuka, M. *et al.* Hypersusceptibility to vesicular stomatitis virus infection in Dicer1-deficient mice is due to impaired miR24 and miR93 expression. *Immunity* **27**, 123–134 (2007).
8. Lecellier, C. H. *et al.* A cellular microRNA mediates antiviral defense in human cells. *Science* **308**, 557–560 (2005).
9. Shimakami, T. *et al.* Stabilization of hepatitis C virus RNA by an Ago2-miR-122 complex. *Proc. Natl Acad. Sci. USA* **109**, 941–946 (2012).

10. Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science* **309**, 1577–1581 (2005).
11. Griffin, D. E. in *Fields Virology* (ed. Knipe, D. M.) 1023–1068 (Lippincott Williams & Wilkins, 2006).
12. Armstrong, P. M. & Andreadis, T. G. Eastern equine encephalitis virus—old enemy, new threat. *N. Engl. J. Med.* **368**, 1670–1673 (2013).
13. Gardner, C. L. *et al.* Eastern and Venezuelan equine encephalitis viruses differ in their ability to infect dendritic cells and macrophages: impact of altered cell tropism on pathogenesis. *J. Virol.* **82**, 10634–10646 (2008).
14. Gardner, C. L., Ebel, G. D., Ryman, K. D. & Klimstra, W. B. Heparan sulfate binding by natural eastern equine encephalitis viruses promotes neurovirulence. *Proc. Natl Acad. Sci. USA* **108**, 16026–16031 (2011).
15. Silverman, M. A. *et al.* Eastern equine encephalitis in children, Massachusetts and New Hampshire, USA, 1970–2010. *Emerg. Infect. Dis.* **19**, 194–201 (2013).
16. MacDonald, G. H. & Johnston, R. E. Role of dendritic cell targeting in Venezuelan equine encephalitis virus pathogenesis. *J. Virol.* **74**, 914–922 (2000).
17. Enright, A. J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
18. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genet.* **39**, 1278–1284 (2007).
19. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012).
20. Mildner, A. *et al.* Mononuclear phagocyte miRNome analysis identifies miR-142 as critical regulator of murine dendritic cell homeostasis. *Blood* **121**, 1016–1027 (2013).
21. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
22. Bernard, K. A., Klimstra, W. B. & Johnston, R. E. Mutations in the E2 glycoprotein of Venezuelan equine encephalitis virus confer heparan sulfate interaction, low morbidity, and rapid clearance from blood of mice. *Virology* **276**, 93–103 (2000).
23. Langlois, R. A., Varble, A., Chua, M. A., Garcia-Sastre, A. & tenOever, B. R. Hematopoietic-specific targeting of influenza A virus reveals replication requirements for induction of antiviral immune responses. *Proc. Natl Acad. Sci. USA* **109**, 12117–12122 (2012).
24. Scott, T. W. & Weaver, S. C. Eastern equine encephalomyelitis virus: epidemiology and evolution of mosquito transmission. *Adv. Virus Res.* **37**, 277–328 (1989).
25. Tesfay, M. Z. *et al.* Alpha/beta interferon inhibits cap-dependent translation of viral but not cellular mRNA by a PKR-independent mechanism. *J. Virol.* **82**, 2620–2630 (2008).
26. Bick, M. J. *et al.* Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562 (2003).
27. Anishchenko, M. *et al.* Generation and characterization of closely related epizootic and enzootic infectious cDNA clones for studying interferon sensitivity and emergence mechanisms of Venezuelan equine encephalitis virus. *J. Virol.* **78**, 1–8 (2004).
28. Aguilar, P. V. *et al.* Structural and nonstructural protein genome regions of eastern equine encephalitis virus are determinants of interferon sensitivity and murine virulence. *J. Virol.* **82**, 4920–4930 (2008).
29. Thomas, J. M., Klimstra, W. B., Ryman, K. D. & Heidner, H. W. Sindbis virus vectors designed to express a foreign protein as a cleavable component of the viral structural polyprotein. *J. Virol.* **77**, 5598–5606 (2003).
30. Brault, A. C. *et al.* Venezuelan equine encephalitis emergence: enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. *Proc. Natl Acad. Sci. USA* **101**, 11344–11349 (2004).

Acknowledgements We thank M. Dunn for excellent technical support, and M. Diamond for reading of the manuscript. This work was supported by National Institutes of Health (NIH) training grants AI049820-10 and AI060525-08 (D.W.T.), research grants AI083383 and AI095436 (W.B.K.) and a Project Grant (K.D.R.) from National Institute of Allergy and Infectious Diseases through the Pacific Northwest Regional Centers for Excellence in Biodefense and Emerging Infectious Diseases Research (U54 AI081680). The views expressed are those of the authors and do not necessarily represent the views of the NIH.

Author Contributions D.W.T., S.C.W., K.D.R. and W.B.K. designed the experiments and analysed the data. D.W.T., C.L.G., A.D.H. and E.W. performed the experiments. C.S., E.C. and A.M. provided key reagents. D.W.T. and W.B.K. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.B.K. (klimstra@pitt.edu).

METHODS

Cell culture. Baby hamster kidney cells (BHK-21), L929 fibroblasts and RAW 264.7 (RAW) macrophage/monocyte cells were maintained as described previously¹³. BMDCs from CD-1, *Ifnar1*^{-/-} and *Mir142*^{-/-} mice²⁰ were generated and maintained as described previously¹³. However, *Mir142*^{-/-} bone marrow was collected and frozen in 10% dimethyl sulphoxide and 90% FBS before culture. *Aedes albopictus* C6/36 mosquito cells were maintained in minimum essential medium alpha medium (Cellgro) supplemented with 10% FBS and 1% L-glutamine (Gibco).

Translation reporters and dual luciferase assay. Wild-type EEEV, wild-type VEEV and host mimic luciferase-expressing translation reporters (diagram in Extended Data Fig. 1) were generated previously^{25,26}. The host mimic translation reporters encode a short 5' and 3' NTR fused in-frame with the fLuc gene. The EEEV 5' ΔNTR and VEEV 5' ΔNTR reporters were constructed using QuikChange II XL mutagenesis kit (Agilent Technologies) and the primers listed in Extended Data Table 1. The chimaeric 5' host 3' EEEV and 5' host 3' VEEV translation reporter was constructed by addition of a restriction endonuclease site into the host fLuc mimic reporter using QuikChange II XL mutagenesis and the primers listed in Extended Data Table 1. The 3' NTR of EEEV or VEEV was then placed into the host fLuc mimic after the fLuc gene with endonuclease digestion. Translation assays were performed as described previously with modifications¹³. Each *in vitro*-transcribed reporter RNA (7.5 μg per reaction) and the *Renilla* reporter RNA (0.75 μg per reaction) were electroporated into RAW and BHK-21 cells (6 × 10⁶ cells per reaction) using the Neon Transfection system (Invitrogen; BHK-21: 1,200 V, 30 ms, 1 pulse; RAW: 1750 V, 25 ms, 1 pulse). Two reactions were combined and aliquoted in triplicate per time point per experiment. Firefly relative light units (RLU) data were normalized to *Renilla* RLUs in each sample.

Viruses. Construction of cDNA clones of VEEV ZPC 738 (ref. 27) (wild-type VEEV), EEEV FL93-939 (ref. 28) (wild-type EEEV) and EEEV 71-77 (ref. 14) were described previously. mCherry, eGFP and nLuc (Promega) reporter viruses were constructed as a cleavable in-frame fusion between the capsid and E3 proteins using QuikChange II XL mutagenesis kit with the first 5 amino acids of E3 fused in-frame to the amino terminus of the reporter genes and the 2A-like protease of *Thosaea asigna* virus (TaV) fused to the carboxy terminus^{29,31}. The EEEV 11337, 142del and 142pm mutants were generated using the EEEV FL93-939 complementary DNA clone and QuikChange mutagenesis II XL kit and primers listed in Extended Data Table 2. Wild-type EEEV nsP3-nLuc translation reporter virus was constructed with the nLuc gene fused in frame with nsP3^{13,31}. The 3' NTR of each mutant was placed into TaV and nsP3-nLuc reporter viruses using EcoRI and Not I.

Virus infections and plaque assay. Virus growth curves were performed as described previously using eGFP or mCherry reporter viruses described above¹³. In brief, BHK-21, RAW and C6/36 cells (2 × 10⁵ cells per well) were infected in triplicate in 24-well plates at a multiplicity of infection (m.o.i.) of 0.1 p.f.u. per cell or 1 p.f.u. per cell (C6/36). BMDCs (1 × 10⁵ cells) were infected in triplicate at a m.o.i. of 5 p.f.u. per cell in suspension, washed and transferred into 24-well plates. For growth curves, supernatant was collected at time zero and indicated time points for titration by plaque assay on BHK-21 cells. For nsP3 translation reporter assays, RAW (4 × 10⁵ cells per well) were infected at an m.o.i. of 1 p.f.u. per cell in triplicate with nsP3-nLuc reporter viruses and collected at indicated time points using 1× Passive Lysis Buffer (PLB; Promega). nLuc expression was quantified using the Nano-Glo Luciferase assay system (Promega) according to manufacturer's guidelines and normalized to protein concentration using a Pierce BCA protein assay (Thermo Scientific).

miRNA overexpression. The pCMV-miR-142 expression plasmid and a neuron-specific miRNA, pCMV-miR-124, (OriGene) were electroporated (4 μg each) into BHK-21 cells (1 × 10⁶ cells) using the Amaxa Nucleofector Kit L (Lonza) according to manufacturer's guidelines. After ~18 h, cells were infected with wild-type EEEV, wild-type VEEV or EEEV 11337 mCherry TaV viruses (m.o.i. = 1 p.f.u. per cell) for 8 h followed by fixation. Co-expression of the miRNA (eGFP) and virus reporter (mCherry) was determined using fluorescence microscopy as described previously¹³. Data are represented as the ratio of the percentage of cells co-expressing the microRNA, miR-142, (eGFP) and virus-infected (mCherry) to cells co-expressing miR-124 and virus-infected.

RT-PCR to detect miR-142-3p. Total cellular RNA was collected from cells using Trizol (Life Technologies) and 1-bromo-3-chloropropane and isopropanol. RNA (200 ng) was reverse transcribed with the miScript II RT kit (Qiagen) according to

manufacturer's guidelines. cDNA was diluted in water and mature miR-142-3p was quantified using the miScript SYBR Green PCR kit (Qiagen) according to manufacturer's guidelines using the miR-142-3p-specific primer 5'-TGTAAGTGT TTCCTACTTTATGGA-3'. miR-142-3p expression was normalized to RNU6B using the primer 5'-GATGACACGCAATTCGTGA-3' and the ΔΔC_T method. Data are calculated as fold change in expression compared to expression of miR-142-3p in BHK-21 cells in which miR-142-3p expression was undetected.

Mouse infections, tissue collection, nLuc analysis. Six-week-old female outbred CD-1 mice (Charles River Laboratories) and 6–9-week-old female or male *Ifnar1*^{-/-} mice bred in house were randomly distributed, infected subcutaneously in both footpads and scored daily for clinical signs and weight loss as described previously^{13,14}. Two investigators were used to analyse the clinical symptoms observed in the mice during morbidity and mortality studies. Investigators were not privy to timing and onset of clinical symptoms from previous experiment. For tissue collection, mice were infected with 10⁵ p.f.u. of nLuc- or mCherry-reporter TaV viruses. PLNs were collected at indicated time points and either frozen on dry ice, placed in 200 μl 1× PLB for nLuc analysis or placed in 4% paraformaldehyde (PFA) for 1 week for visualization on a fluorescence microscope. PLNs were homogenized and analysed for nLuc expression using the Nano-Glo Luciferase assay system. All EEEV PLNs were photographed using equal exposure times (615 ms) whereas the VEEV lymph node, owing to an increased signal, was visualized using a lower exposure time (90 ms) at ×4 magnification using cellSens Standard software (Olympus). The 71-77/11337 inset was imaged at ×40 magnification and 457-ms exposure length. Brightness and contrast of the images were adjusted equally using Adobe Photoshop CS3 and Microsoft Powerpoint software. All animal procedures were carried out in accordance with American Association for the Accreditation of Laboratory Animal Care International-approved institutional guidelines for animal care and use and approved by the University of Pittsburgh Institutional Animal Care and Use Committee. No statistical methods were used to ensure adequate power. Sample sizes were chosen based upon experience with the mortality kinetics of EEEV in mice and historical group number requirements to achieve statistical significance yet use the fewest animals possible.

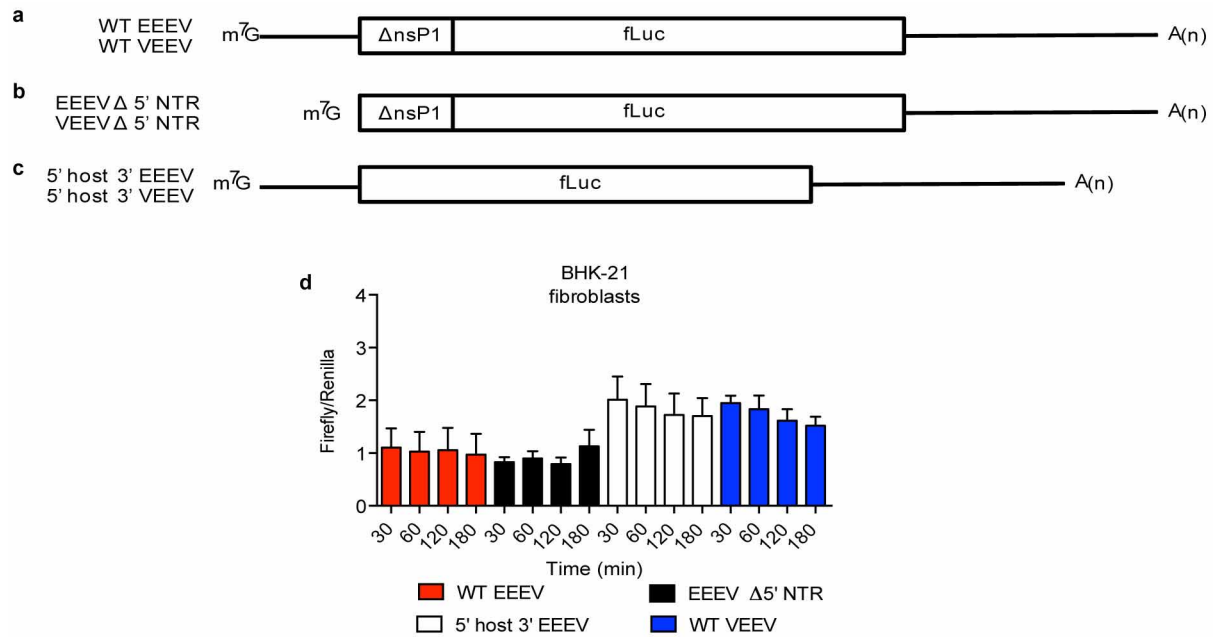
Quantification of virus-infected cells from PLN. To quantify number of virus infected cells in PLNs, CD-1 mice were infected as above, collected at 12 h.p.i., minced and incubated with Liberase TL (0.2 mg ml⁻¹; Roche) and DNase I (0.2 mg ml⁻¹; Roche) for 20 min at 37 °C. After removal of cellular debris, cells were stained with anti-mouse CD16/32 (93; eBioscience), and Fixable Viability Dye eFluor 506 (eBioscience). After washing, cells were stained with anti-mouse CD11b (clone N418, Tonbo Biosciences), and anti-mouse CD11c (M1/70; eBioscience) to identify myeloid cells. Cells were fixed in 4% PFA and analysed using a BD LSRFortessa (BD Bioscience) and FloJo Software (Tree Star). Number of virus-infected cells was calculated based on total number of cells in both PLNs per mouse.

IFN-α/β analysis. Serum IFN-α/β was measured using a standard biological assay on L929 cells as described previously¹³.

Mosquito infection. Adult female *Aedes (Ochlerotatus) taeniorhynchus* mosquitoes were infected with EEEV, 11337, 142del and 142pm mCherry- or eGFP-reporter TaV viruses in artificial bloodmeals. Engorged females were incubated at 27 °C for 10 days under 12 h light/12 h dark circadian lighting conditions and assayed for infection by inoculation onto Vero cell monolayers and observation for cytopathic effects as described previously³⁰.

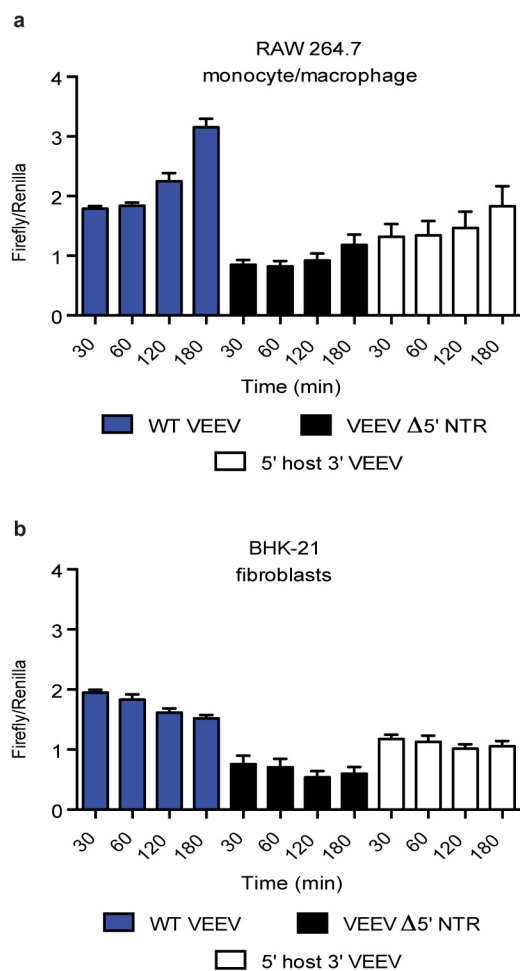
Statistical analysis. Statistical significance for mortality curves was determined by Mantel-Cox log-rank test. For all viral growth curve experiments, data was log₁₀ transformed and unpaired *t*-tests were performed and corrected for multiple comparisons using the Holm-Sidak method with alpha = 0.05. For all other experiments, a two-tailed unpaired *t*-test was used (GraphPad Prism software). Statistical analysis for nLuc expression in CD-1 PLN was performed comparing wild-type EEEV versus the other EEE viruses whereas wild-type VEEV was compared to only 71-77/11337 at each time point. nLuc expression in *Ifnar1*^{-/-} mice was compared to the corresponding viruses at 12 h.p.i. in CD-1 mice only. Statistical analysis for quantification of the number of infected cells in PLN was performed between the EEEV mutant viruses and wild-type EEEV, and between wild-type VEEV and 71-77/11337.

31. Sun, C., Gardner, C. L., Watson, A. M., Ryman, K. D. & Klimstra, W. B. Stable, high-level expression of reporter proteins from improved alphavirus expression vectors to track replication and dissemination during encephalitic and arthritogenic disease. *J. Virol.* (<http://dx.doi.org/10.1128/JVI.02990-13>) (4 December 2013).

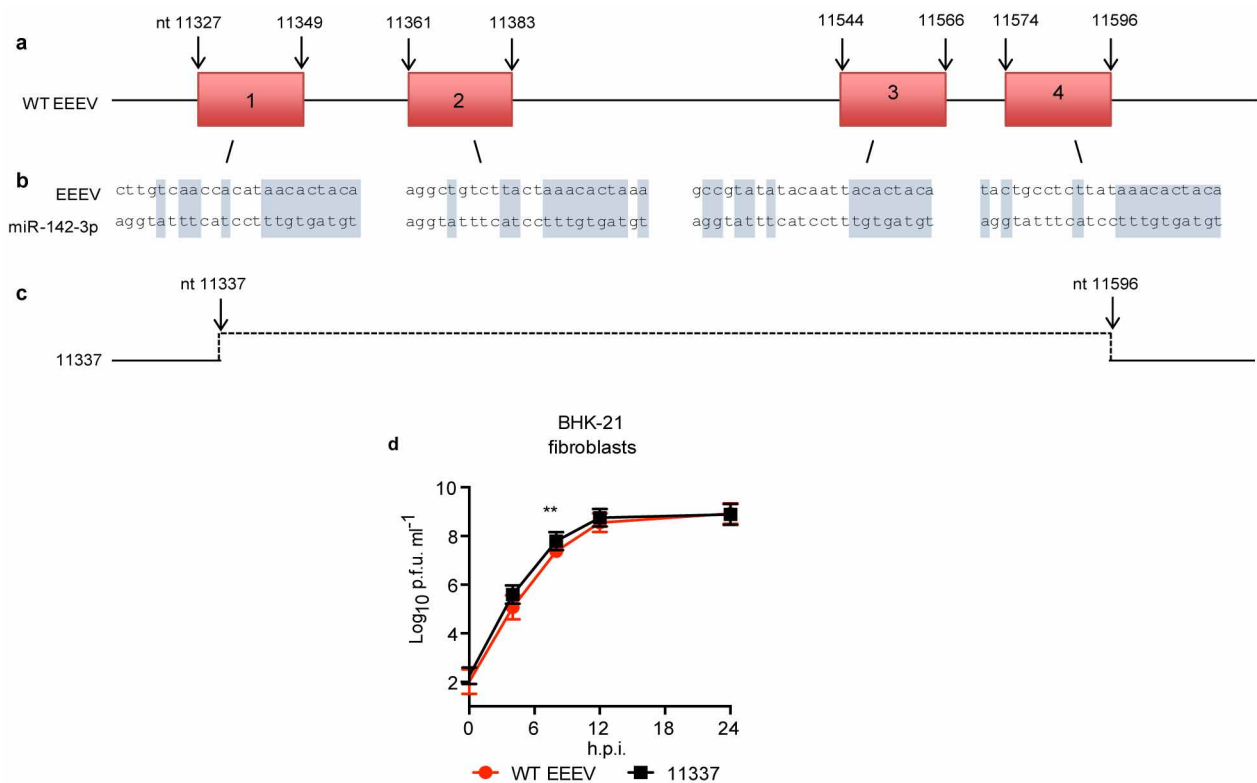


Extended Data Figure 1 | EEEV 3' NTR does not restrict translation in BHK-21 fibroblasts. **a**, Wild-type EEEV and wild-type VEEV translation reporters encode the translational initiation control sequences fused to the Firefly luciferase (fLuc) gene. **b**, The EEEV 5' Δ NTR and VEEV 5' Δ NTR encode the truncated nsP1 gene and only the 3' NTR of either EEEV or VEEV, respectively. **c**, The 3' NTR of EEEV or VEEV was inserted into a host mRNA

mimic reporter to generate the 5' host 3' EEEV or 5' host 3' VEEV reporters. All translation reporters contain a 5' cap and a 3' poly(A) tail. **d**, Translation of wild-type EEEV, EEEV 5' Δ NTR, and 5' host 3' EEEV reporters in BHK-21 cells. Error bars represent mean \pm s.d. and the data are averaged from three independent experiments performed in triplicate. WT, wild type.

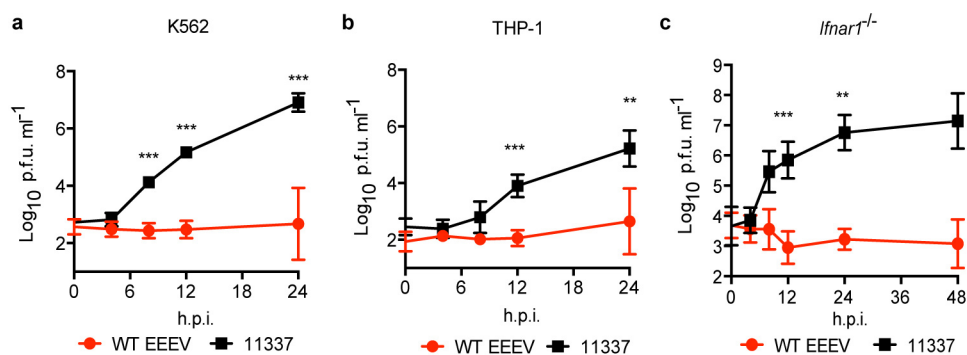


Extended Data Figure 2 | VEEV 3' NTR does not restrict translation in myeloid cells. a, b, Translation of wild-type VEEV, VEEV 5' Δ NTR, and 5' host 3' VEEV reporters in RAW (a) and BHK-21 (b) cells. Error bars represent mean \pm s.d. and the data are averaged from three independent experiments performed in triplicate.



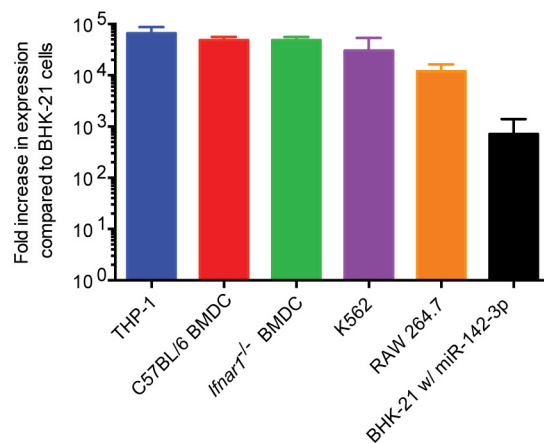
Extended Data Figure 3 | Removal of miR-142-3p binding sites in the 3' NTR of EEEV does not alter replication in BHK-21 fibroblasts. **a**, Red boxes indicate the four miR-142-3p binding sites in the 3' NTR. Numbers represent nucleotide (nt) positions at the start and end of each miRNA binding site. **b**, Grey boxes correspond to the complementary nts in the EEEV 3' NTR

and miR-142-3p. **c**, EEEV mutant 11337 contains a deletion in the 3' NTR from nt 11,337 to 11,596. **d**, Replication of wild-type EEEV and 11337 in BHK-21 cells. $n = 3$ independent experiments. Error bars indicate geometric mean \pm s.d., and asterisks indicate differences that are statistically significant (** $P < 0.01$).

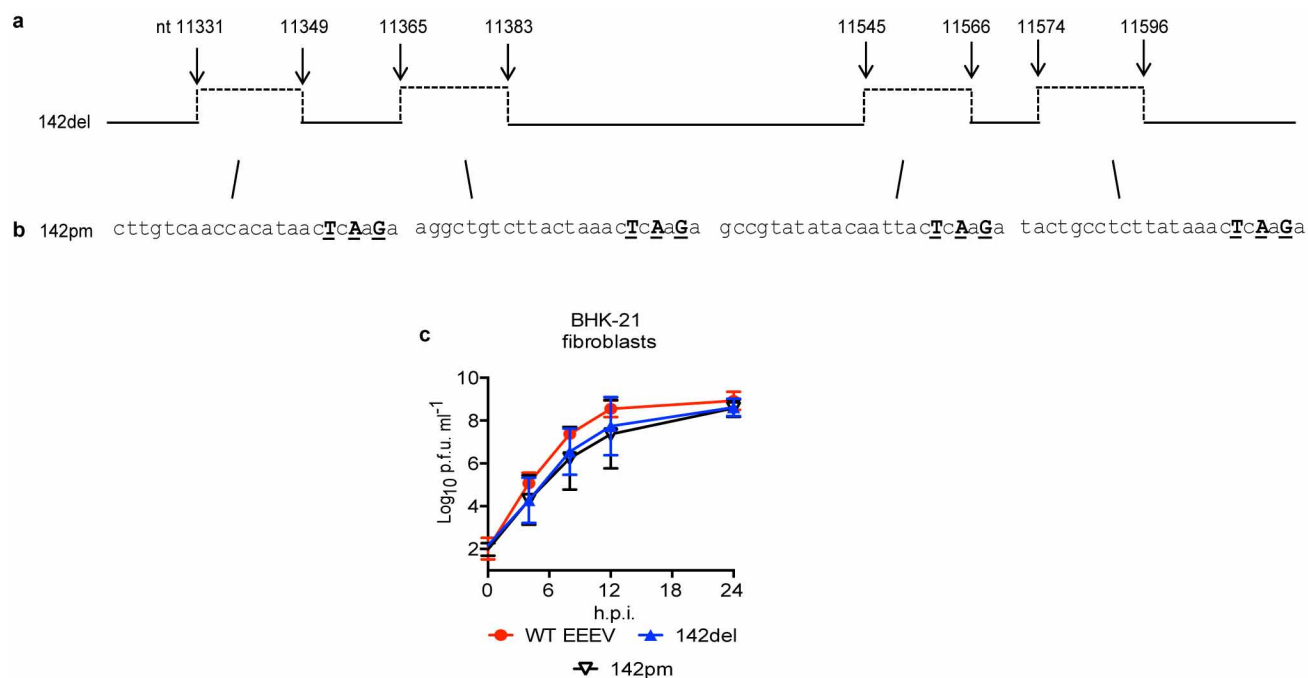


Extended Data Figure 4 | miR-142-3p binding sites in EEEV restrict replication in human macrophage/monocyte cell lines and primary murine *Ifnar1*^{-/-} BMDCs. **a, b**, Replication of wild-type EEEV and 11337 in human K562 (**a**) and THP-1 (**b**) cells. *n* = 2 (THP-1) and 3 (K562) independent experiments. **c**, Removal of type I IFN does not alleviate wild-type EEEV

restriction in primary murine *Ifnar1*^{-/-} BMDCs. *n* = three independent experiments. Data represent the geometric mean \pm s.d., and asterisks indicate differences that are statistically significant (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

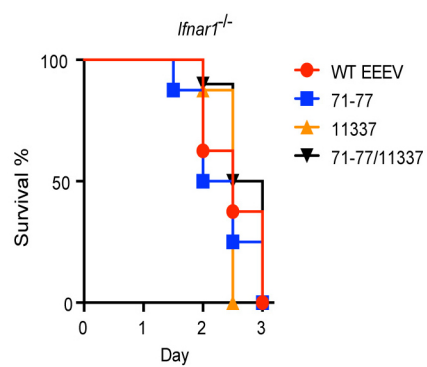


Extended Data Figure 5 | Relative expression of miR-142-3p in mouse and human cells. Quantitative RT-PCR on primary murine BMDCs, murine and human monocyte/macrophage cell lines, and BHK-21 cells expressing miR-142-3p (BHK-21 w/miR-142-3p). Fold increase in expression is calculated compared to expression of miR-142-3p in BHK-21 cells in which miR-142-3p expression was undetectable.

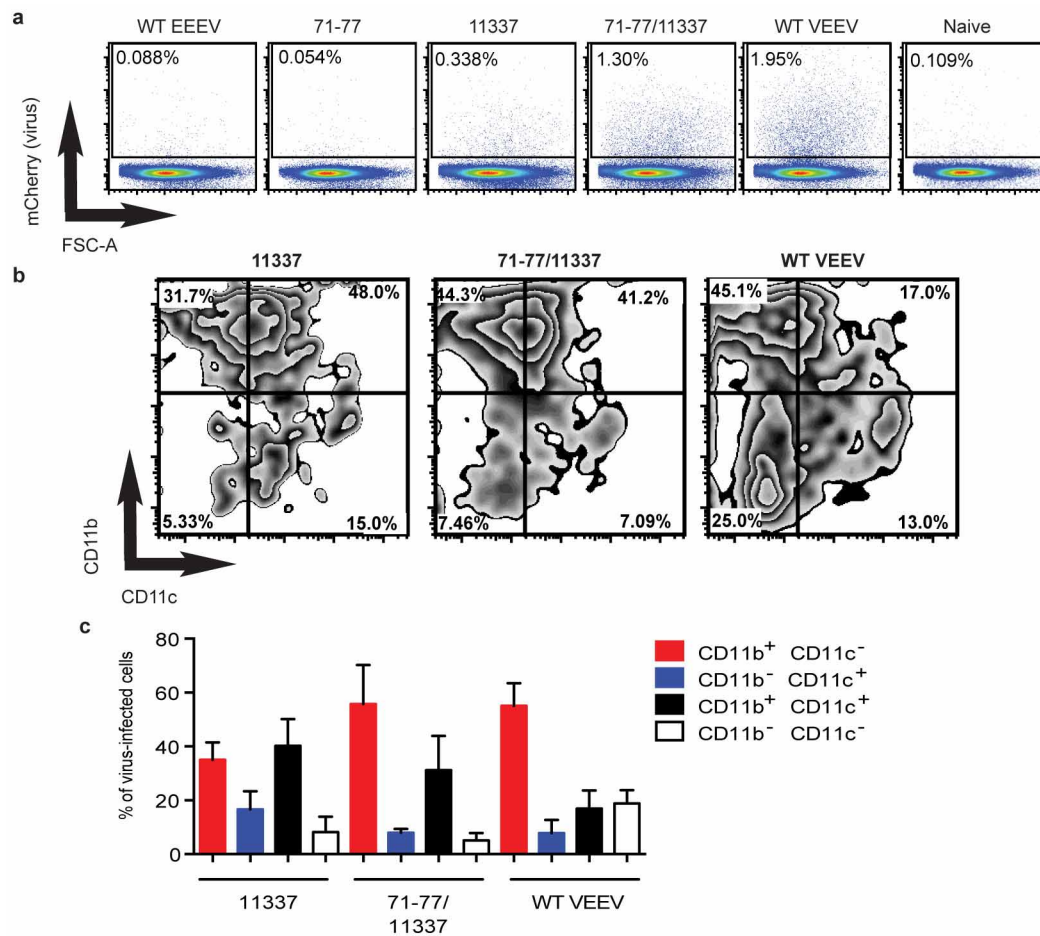


Extended Data Figure 6 | Specific deletion of the miR-142-3p binding sites in the 3' NTR of wild-type EEEV does not alter replication in BHK-21 fibroblasts. **a**, EEEV 142del virus contains four deletions corresponding to the complementary nucleotides in the 3' NTR that bind to miR-142-3p, eliminating all four miR-142-3p binding sites. **b**, EEEV 142pm virus contains three point

mutations in each of the miR-142-3p binding sites that correspond to the seed sequence of miR-142-3p. **c**, Replication of wild-type EEEV and 11337 in BHK-21 cells. $n =$ three independent experiments. Error bars indicate geometric mean \pm s.d.



Extended Data Figure 7 | Type I IFN attenuates 11337 and 71-77/11337.
Survival curves in *Ifnar1*^{-/-} mice. $n = 8$ and 10 (71-77/11337) mice per virus from two independent experiments.



Extended Data Figure 8 | EEEV 11337 and 71-77/11337 infect myeloid lineage cells in the PLN. **a**, Per cent virus-infected cells in PLN in naïve, wild-type VEEV-, wild-type EEEV-, 71-77-, 11337- and 71-77/11337-infected mice. Plots are representative of $n = 4$ (naïve), 5 (71-77) or 6 (71-77/11337) mice from two independent experiments. **b**, **c**, Wild-type VEEV, 11337 and 71-77/11337 infect myeloid lineage cells in the PLN. **b**, Representative flow plot from 1

mouse of CD11b (y-axis) and CD11c (x-axis) expression on virus-infected cells. $n = 4$ (naïve), 5 (71-77) or 6 (71-77/11337) mice from two independent experiments. **c**, Summary of CD11b and CD11c expression on virus-infected cells from wild-type VEEV-, 11337- and 71-77/11337-infected PLNs. Only mice with responses above naïve mice background levels were used to determine CD11b and CD11c expression.

Extended Data Table 1 | Primers used to generate the translation reporters using the QuikChange II XL mutagenesis kit

Reporter	Primer Name	Sequence
EEEV Δ5' NTR	EEEV Δ5' NTR-F	5'-agctcggatcctaatacgaactcactatagatggagaaagttcatgttgacttagacgca-3'
	EEEV Δ5' NTR-R	5'-tgcgctctaagtcaacatgaactttctccatctatagtgagtcgtattaggatccgagct-3'
VEEV Δ5' NTR	VEEV Δ5' NTR-F	5'-agaggatccctaatacgaactcactatagatggagaaagttcacgttgacatcgaggaa-3'
	VEEV Δ5' NTR-R	5'-ttcctcgatgtcaacgtgaactttctccatctatagtgagtcgtattaggatcctct-3'
Host mimic w/Not1	fLuc-Not1-F	5'-ccccaaaaaaaaaaaaaaaaaaaaaaaaaagcgccgcgtaatcatgtcatagc-3'
	fLuc-Not1-R	5'-gctatgacatgattacggcgccgctttttttttttttttttttttttttttttttggg-3'

Extended Data Table 2 | Primers used in the generation of EEEV mutant viruses

Virus	Primer Name	Sequence
11337	EEEV11337-F	5'-gacattaacatcttgtcaaccggcagcgcataatgctgtcttttatatc-3'
	EEEV11337-R	5'-gatataaaagacagcattatgcgctgccggttgacaagatgttaatgtc-3'
142del	EEEV-Del-1-2-F	5'-catagacattaacatcttgggcagtgtataaggcttcaccctagttcgatgtacttccg-3'
	EEEV-Del-1-2-R	5'-cggaagtacatcgaactaggggtgaagccttatacactgcccaagatgttaatgtctatg-3'
	EEEV-Del-3-4-F	5'-ctttataatcaggcataattgggtaataataccgcctggcagcgcataatgctgtc-3'
	EEEV-Del-3-4-R	5'-gacagcattatgcgctgccagcggtatattaccaattatgcctgattataaa-3'
142pm	Δ miR-142-1-F	5'-taacatcttgtcaaccacataactcaagaggcagtgt-3'
	Δ miR-142-1-R	5'-attgtagaacagttgggtattgagttctccgtcacat-3'
	Δ miR-142-2-F	5'-taaggctgtcttactaaactcaagattcacctag-3'
	Δ miR-142-2-R	5'-ctaggggtgaatcttgagtttagtaagacagcctta-3'
	Δ miR-142-3-F	5'-gcataattgccgtatatacaattactcaagaggtaataataccgcctcttataaa-3'
	Δ miR-142-3-R	5'-tttataagaggcggtatattacctcttgagtaattgtatatacggcaattatgc-3'
	Δ miR-142-4-F	5'-ataccgcctcttataaaactcaagaggcagcgc-3'
	Δ miR-142-4-R	5'-gcgctgcctcttgagtttataagaggcggtat-3'

RecA bundles mediate homology pairing between distant sisters during DNA break repair

Christian Lesterlin¹, Graeme Ball¹, Lothar Schermelleh¹ & David J. Sherratt¹

DNA double-strand break (DSB) repair by homologous recombination has evolved to maintain genetic integrity in all organisms¹. Although many reactions that occur during homologous recombination are known^{1–3}, it is unclear where, when and how they occur in cells. Here, by using conventional and super-resolution microscopy, we describe the progression of DSB repair in live *Escherichia coli*. Specifically, we investigate whether homologous recombination can occur efficiently between distant sister loci that have segregated to opposite halves of an *E. coli* cell. We show that a site-specific DSB in one sister can be repaired efficiently using distant sister homology. After RecBCD processing of the DSB, RecA is recruited to the cut locus, where it nucleates into a bundle that contains many more RecA molecules than can associate with the two single-stranded DNA regions that form at the DSB. Mature bundles extend along the long axis of the cell, in the space between the bulk nucleoid and the inner membrane. Bundle formation is followed by pairing, in which the two ends of the cut locus relocate at the periphery of the nucleoid and together move rapidly towards the homology of the uncut sister. After sister locus pairing, RecA bundles disassemble and proteins that act late in homologous recombination are recruited to give viable recombinants 1–2-generation-time equivalents after formation of the initial DSB. Mutated RecA proteins that do not form bundles are defective in sister pairing and in DSB-induced repair. This work reveals an unanticipated role of RecA bundles in channelling the movement of the DNA DSB ends, thereby facilitating the long-range homology search that occurs before the strand invasion and transfer reactions.

Although repair of DSBs by homologous recombination is efficient in *E. coli*, it is not known whether homologous recombination events are restricted to the ~15 min period in which sister loci remain in close association after replication^{4,5}, or whether homologous recombination can occur between distant segregated sister loci (~1.3 μm apart). To test whether segregated distant sisters can undergo recombination when a site-specific DSB was introduced into one sister ~1 megabase pairs (Mb) from the replication *oriC*, we expressed small amounts of unstable I-SceI endonuclease⁶ so that most cells had only one of two sisters cleaved, and repaired the DSB efficiently in reactions that required RecABCD (Fig. 1 and Extended Data Fig. 1).

Fluorescent ParB protein binding to DNA on either side of the cut site allowed us to monitor the fate of cut and uncut loci. After DSB induction, loss and randomized positioning of foci occurred when compared to the uncut control, in which one-focus cells had the focus displaced slightly from mid-cell and two-focus cells had the foci on either side of mid-cell. Time-lapse analysis showed focus pairing, accompanied by increased fluorescence and focus loss dependent on RecBCD degradation; events absent in wild-type cells (Fig. 1A, C and Extended Data Figs 1a, 2). RecABCD-dependent sister pairing occurred on average ~100 min after induction of I-SceI expression and persisted for ~50 min before sister re-segregation (Fig. 1A). The two ends of the DSB remained in proximity after cutting and during subsequent processing, supporting schemes in which both DNA ends are engaged in homologous recombination repair (Fig. 1C and Extended Data Fig. 2A), as reported for eukaryotic repair⁷. The behaviour of fluorescent foci marking each of

the cut ends was similar, with the foci moving to the cell periphery after cutting (compare Fig. 1A and Extended Data Fig. 1b, d). RecBCD processing of the DSB was bidirectional, consistent with efficient loading of RecBCD complexes onto both ends of the break. The rate of DNA degradation in the absence of RecA was ~190 bp s⁻¹ per RecBCD complex, eventually leading to complete nucleoid degradation (Fig. 1C, D, Extended Data Fig. 2B, C, Supplementary Video 1 and Supplementary Discussion).

RecBCD-dependent recruitment of RecA to the DSB was assayed in cells expressing a carboxy-terminal fusion of RecA to green fluorescent protein (GFP)⁸ and wild-type RecA, both expressed from wild-type chromosomal *recA* promoters. These cells exhibited wild-type repair and recombination, whereas those expressing RecA–GFP alone were not fully repair proficient⁸ (Extended Data Fig. 3A). Before DSB induction, ~95% of cells showed RecA–GFP fluorescence that was uniformly distributed throughout the cell, with ~5% of cells having fluorescent spots that were not associated with a marked locus (21% colocalization) (Extended Data Fig. 3B). After DSB induction, fluorescent RecA spots appeared close to or coincident with one of the two marked sister loci (64% colocalization; Fig. 2a, b). The transient RecA spots nucleated rapidly into filamentous structures that we term RecA bundles, which had formed their maximum length by ~13 min and most often extended along the cell (Fig. 2c, d and Supplementary Video 2). The DSB-induced stimulation of RecA spot and bundle formation required RecBCD processing of the cut ends (Fig. 2e).

Rapid sister locus pairing occurred ~47 min after RecA bundle formation (Fig. 2f, g). During this period, no consistent changes in bundle architecture occurred and no significant turnover of RecA within the mature bundle structure was observed by fluorescence recovery after photobleaching (FRAP; Extended Data Fig. 3C and Supplementary Video 3). Once initiated, sister focus pairing was rapid (within 5 min in 69% of events) and was asymmetric, with the cut sister travelling ~84% of the ~1.3 μm distance between sister loci (Fig. 2h). Although the cut and uncut loci had similar diffusion coefficients (D_{app}), the movement of the cut locus was 5–7-fold more directional than the uncut locus, indicative of channelling of the movement of the cut locus (Fig. 2h and Extended Data Fig. 4A).

We propose that nucleation of RecA onto the single-stranded (ss)DNA ends formed by RecBCD processing at the DSB triggers polymerization of RecA into bundles, which are required for pairing between distant sisters. RecA bundles disassembled rapidly and unidirectionally ~17 min after pairing of the sister loci, possibly triggered by the loss of the ssDNA–RecA ends upon strand invasion (Fig. 2i, j and Supplementary Video 2). After pairing, sisters re-segregated and cells re-established normal growth, cell division and chromosome segregation (Fig. 2k, summary diagram). In parallel, we have shown that DSB induction and bundle formation are followed by the recruitment of PriA and RecG recombination proteins, which promote replication restart and recombination intermediate processing. By contrast, RecN was recruited early to the break (Extended Data Fig. 4B, C and Supplementary Discussion). Bundles of wild-type RecA that formed in response to DSBs were also observed in fixed cells using immunocytochemistry and in live cells with a more functional

¹Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

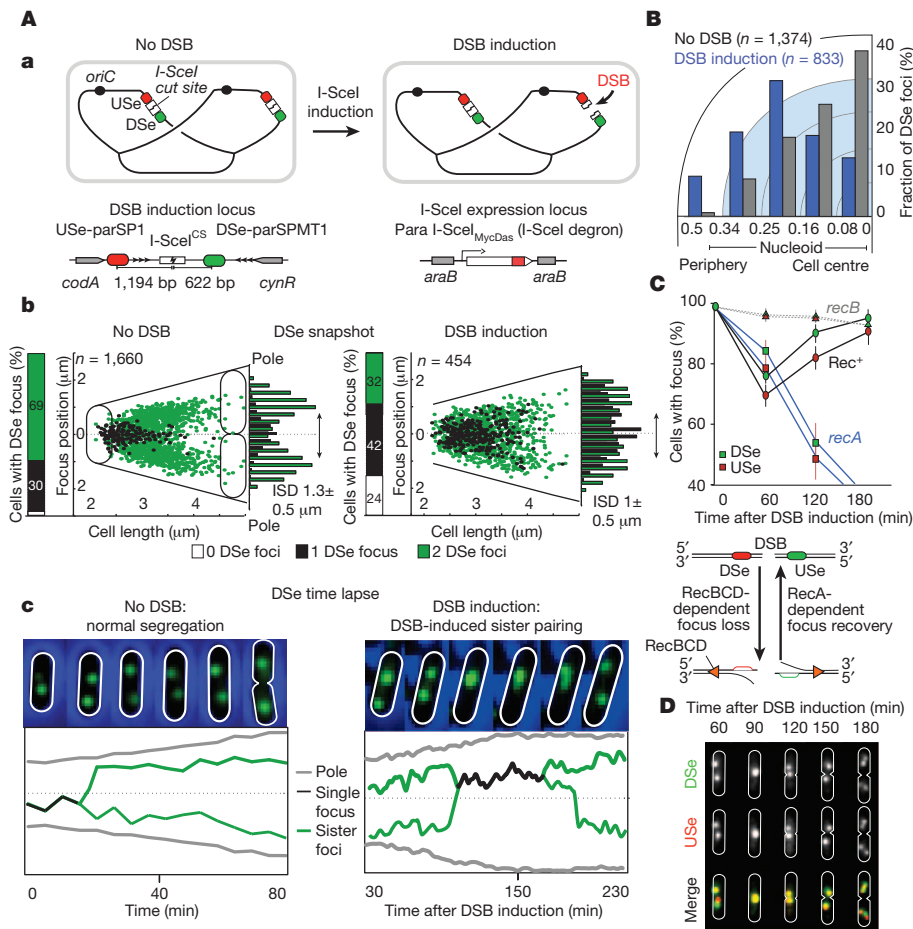


Figure 1 | Effects of DSB induction on positioning of adjacent chromosome loci. **A**, **a**, schematic of DSB induction. **b**, Snapshot analysis of downstream end (DSe) focus localization before (left) and after (right) 60 min of DSB induction. ISD, mean inter-sister distance. **c**, Time-lapse analysis (5 min per frame) of DSe focus dynamics. **B**, Distributions of DSe foci along the cell diameter before and after 60 min of DSB induction. Data are shown for cut sites that changed

position after DSB induction. **C**, DSB-induced reduction of the fraction of cells with DSe and upstream-end (USE) foci is RecBCD dependent and indicates simultaneous processing of both DSB ends. Error bars indicate standard deviations. **D**, DSe and USE foci remain coincident during sister pairing and subsequent sister separation.

fusion between RecA803 and red fluorescent protein (RFP)⁹ (Extended Data Fig. 5). Furthermore, repair-deficient cells expressing RecA(K72A)–GFP, which fails to bind ATP, did not form spots or bundles, whereas cells expressing RecA(K72R)–GFP, which binds ATP but is hydrolysis impaired¹⁰, formed weak spots close to the cut locus ($0.49 \pm 0.2 \mu\text{m}$), consistent with RecA(K72R) forming filaments on ssDNA and initiating strand transfer *in vitro*¹⁰. Nevertheless, the RecA(K72R)–GFP spots failed to nucleate bundles, and *recA*^{K72R} cells did not show sister pairing and were repair deficient (Extended Data Fig. 5). These data demonstrate that RecA bundles are physiologically relevant forms of RecA that are critical for homologous recombination repair between distant regions of homology.

To gain more insight into RecA bundle architecture and function, we used super-resolution three-dimensional structured illumination microscopy (3D-SIM) and live-cell time-lapse imaging (Fig. 3). RecA bundles most frequently adopted overall longitudinal trajectories, with some radial writhing (Supplementary Video 4) and were located between the surface of the nucleoid and the inner membrane. Bundles may be excluded from the nucleoid because of their size, as are $\sim 20\text{-nm}$ -diameter ribosomes¹¹, and their localization at the cell membrane could arise because of association of activated forms of RecA with phospholipid membranes¹². 3D-SIM showed that 91% of bundles were associated with at least one sister focus, confirming the earlier data (Fig. 3A, Extended Data Fig. 3B and Supplementary Video 5). Bundles were on average $\sim 1.4\text{-}\mu\text{m}$ long and typically consisted of a $\sim 160\text{-nm}$ -thick central

region bounded by thinner ($\leq 120\text{ nm}$) extensions (Extended Data Fig. 3B). 3D-SIM time-lapse analysis revealed that the central region was relatively immobile whereas the thin extensions moved rapidly and appeared to probe the inner cell compartment (Fig. 3A and Supplementary Videos 6, 7).

Bundles contained $\sim 70\%$ of the normal pool of cellular RecA, estimated to be $\sim 4,640 \pm 1,908$ molecules per cell, independently of the presence of spots or bundles (Fig. 3B, C, Extended Data Fig. 3D and Supplementary Discussion). The rate of RecA incorporation into DSB-induced bundles was ~ 4 monomers per second, which compares with an *in vitro* rate of $0.5\text{--}7$ monomers per second per DNA end or nucleation point^{13,14}. We estimate that $>85\%$ of RecA molecules present in bundles are likely to be DNA free and additional to the two presynaptic ssDNA–RecA filaments that we propose nucleate bundle formation (Supplementary Discussion). DNA-free RecA filaments and bundles can form *in vitro*^{15,16}, and have a similar architecture to that of RecA–DNA filaments^{17–19}. The observed $\sim 1.4\text{-}\mu\text{m}$ length of RecA bundles is less than the $\sim 2.5\text{-}\mu\text{m}$ -long twin filament that would form from $\sim 3,200$ RecA molecules nucleating onto the two ssDNA ends at the break. Thus, RecA bundles may contain lateral filaments, potentially explaining the thicker central regions of the bundles. RecA bundles have been observed in previous light microscopy and electron microscopy studies, and have been shown to appear in response to DNA damage in *E. coli*^{18,20,21}, and in response to DSBs in *Bacillus subtilis*²². In the latter case, spots appeared close to the site of the DSB and extended into

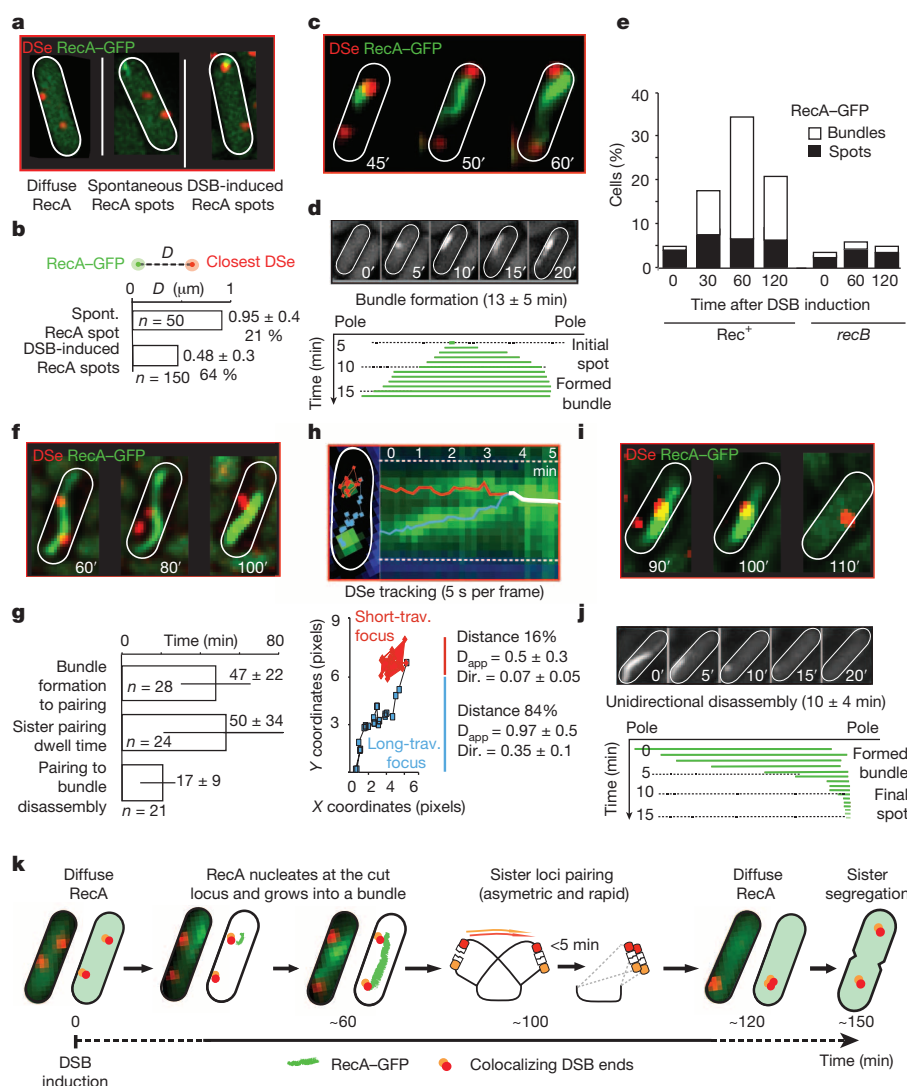


Figure 2 | RecA bundle formation and disassembly, and RecA-mediated sister locus pairing. **a**, RecA-GFP spot formation in relation to the cut locus during sister pairing using wide-field microscopy. **b**, Histograms of the mean distance (D) between the centres of RecA spots and the closest DSe focus with the percentage of colocalization (when $D < 0.5 \mu\text{m}$). Spont., spontaneous. **c**, Wide-field imaging of nucleation of RecA bundles from RecA-GFP spots at the DSe. **d**, Time-lapse images (time in minutes indicated by prime symbol) and analysis of RecA-GFP bundle formation. **e**, Snapshot analysis of RecB-dependent RecA-GFP bundle formation as a function of DSB-induction time (500 cells analysed for each data set). **f**, Wide-field imaging of RecA-GFP

longitudinal bundles, whose formation was dependent on the *B. subtilis* RecBCD orthologue, AddAB. Nevertheless, their physiological significance has been unclear, although it was proposed that RecA bundles are inactive storage forms^{8,18}.

Our data demonstrate conclusively that DSB-induced homologous pairing can occur between distant segregated sister loci, and therefore homologous recombination is not restricted to the ~ 15 min cohesion period after replication. We propose that RecA bundles actively facilitate the process of locus pairing between distant sisters, thereby leading to eventual homology-directed strand invasion. The observation that sister pairing is not necessarily accompanied by bundle shortening suggests that the DNA ends move along the bundle. This could be driven by sliding between RecA filaments within the bundle, as proposed for ParM-mediated DNA movement²³, a treadmilling mechanism in which RecA filament growth at the leading edge and disassembly in the region of the cut locus could lead to rapid pairing, or a process in which a motor-like protein associated with the cut ends moves along the bundle,

akin to other motor proteins moving along cytoskeletal elements. Mechanisms in which pairing arises in part as 'passive' one-dimensional diffusion along a restraining bundle would still require a process that directs the cut locus towards the opposite side of the cell. In any of these scenarios, the channelled movement of the cut ends would be 'blind' to the uncut sister until a local homology search accompanied by strand invasion generated the strand transfer complex whose processing leads to completed recombination products. It follows, therefore, that nucleation of a RecA spot into a bundle should occur irrespective of whether sister homology is present. Consistent with this, RecA bundles formed at similar frequencies regardless of whether sister homology was present (Extended Data Fig. 3E and Supplementary Discussion). The two-step mechanism we propose only needs to be used when sister loci remain in close proximity, DSB repair by homologous recombination might occur simply by the canonical mechanism of strand invasion and transfer.

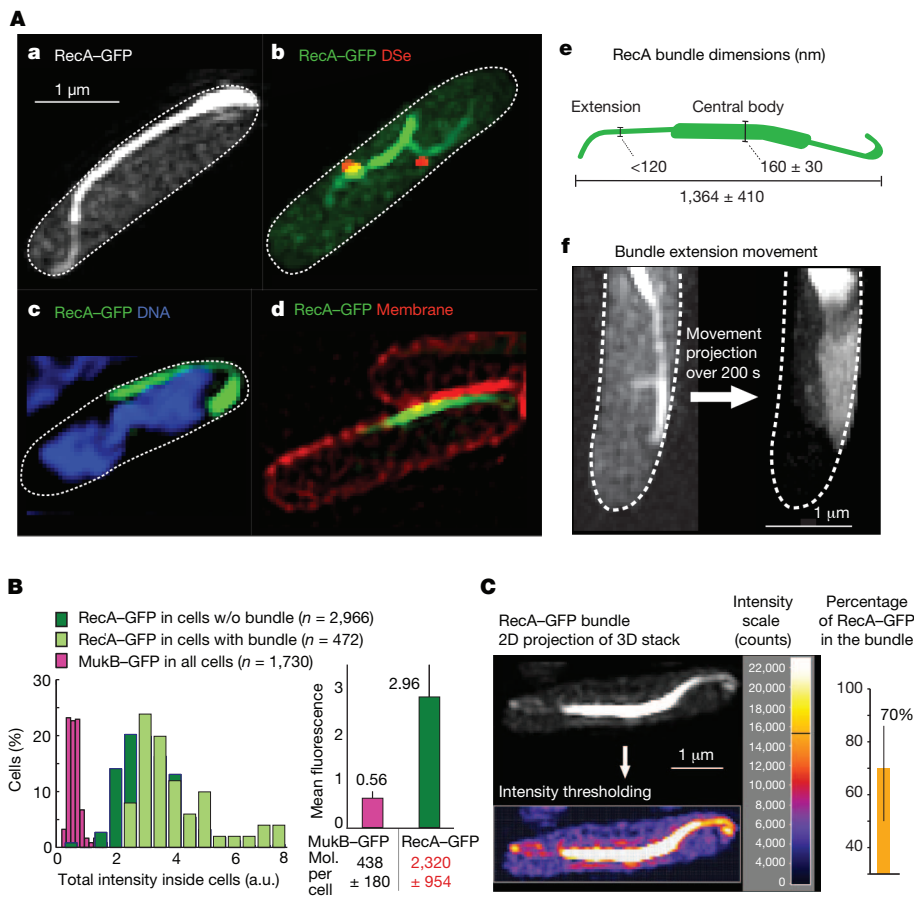


Figure 3 | Super-resolution imaging of RecA bundles reveals their intracellular localization.

A, a–d, 3D-SIM of cells with DSB-induced RecA–GFP bundles alone (**a**), with DSe foci (**b**), relative to DNA (**c**), and relative to membrane (**d**). **e**, Bundle architecture. **f**, The dynamic behaviour of RecA-bundle thin extensions is illustrated by the projection of the movement over 200 s (taken from Supplementary Video 7). **B**, Total fluorescence intensity distribution and average fluorescence of RecA–GFP in cells with or without bundles, and of MukB–GFP. a.u., arbitrary units; Mol., molecules. **C**, The fraction of total cellular RecA monomers within bundles (see Methods). Seventy \pm 18% of the intracellular fluorescence was incorporated into bundles. **B, C**, Error bars indicate standard deviations.

The formation of RecA bundles and equivalent structures in other organisms may be a general mechanism for channelling the movement of damaged DNA loci, thereby enabling a genome-wide homology search in a cell compartment at the periphery of the nucleoid. As other repair proteins are known to interact with RecA and its orthologues, bundles may promote the binding and movement of other players in DNA repair. In eukaryotes, homologous recombination may also fall into two classes, defined by the distance the homology search needs to take. For example, allelic events between cohesed sisters and between paired homologues in meiosis do not involve distant pairing events. By contrast, non-allelic homologous recombination events, or those between homologous mitotic chromosomes, may use processes that involve the equivalent of RecA bundles to facilitate the homology search, a process that may take longer, as observed for homologous recombination between homologues intermediates⁷.

METHODS SUMMARY

Bacteria were derivatives of *E. coli* K-12 TB28 (MG1655; *AlacIZYA*). Mutations were introduced by λ -Red recombination or P1 transduction. Cells were grown at 30 °C in M9 0.2% glucose and antibiotics when required. Exponentially growing cells ($A_{600\text{ nm}}$ 0.1–0.2) were used for microscopy. Full details of bacteria and plasmids used, as well as protocols for sample preparation, imaging and analyses can be found in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 July; accepted 5 November 2013.

Published online 22 December 2013.

1. Cromie, G. A., Connelly, J. C. & Leach, D. R. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell* **8**, 1163–1174 (2001).

2. Cox, M. M. Regulation of bacterial RecA protein function. *Crit. Rev. Biochem. Mol. Biol.* **42**, 41–63 (2007).
3. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642–671 (2008).
4. Lesterlin, C., Gigant, E., Boccard, F. & Espeli, O. Sister chromatid interactions in bacteria revealed by a site-specific recombination assay. *EMBO J.* **31**, 3468–3479 (2012).
5. Reyes-Lamoth, R., Possoz, C., Danilova, O. & Sherratt, D. J. Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell* **133**, 90–102 (2008).
6. Monteilh, C., Perrin, A., Thierry, A., Colleaux, L. & Dujon, B. Purification and characterization of the *in vitro* activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.* **18**, 1407–1413 (1990).
7. Bzymek, M., Thayer, N. H., Oh, S. D., Kleckner, N. & Hunter, N. Double Holliday junctions are intermediates of DNA break repair. *Nature* **464**, 937–941 (2010).
8. Renzette, N. *et al.* Localization of RecA in *Escherichia coli* K-12 using RecA–GFP. *Mol. Microbiol.* **57**, 1074–1085 (2005).
9. Handa, N., Amitani, I., Gumlaw, N., Sandler, S. J. & Kowalczykowski, S. C. Single molecule analysis of a red fluorescent RecA protein reveals a defect in nucleoprotein filament nucleation that relates to its reduced biological functions. *J. Biol. Chem.* **284**, 18664–18673 (2009).
10. Rehauer, W. M. & Kowalczykowski, S. C. Alteration of the nucleoside triphosphate (NTP) catalytic domain within *Escherichia coli* recA protein attenuates NTP hydrolysis but not joint molecule formation. *J. Biol. Chem.* **268**, 1292–1297 (1993).
11. Bakshi, S., Stryaporn, A., Goulian, M. & Weissshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* **85**, 21–38 (2012).
12. Garvey, N., St John, A. C. & Witkin, E. M. Evidence for RecA protein association with the cell membrane and for changes in the levels of major outer membrane proteins in SOS-induced *Escherichia coli* cells. *J. Bacteriol.* **163**, 870–876 (1985).
13. Bell, J. C., Plank, J. L., Dombrowski, C. C. & Kowalczykowski, S. C. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* **491**, 274–278 (2012).
14. Feinstein, E. *et al.* Single-molecule studies of the stringency factors and rates governing the polymerization of RecA on double-stranded DNA. *Nucleic Acids Res.* **39**, 3781–3791 (2011).

15. Brenner, S. L., Zlotnick, A. & Griffith, J. D. RecA protein self-assembly. Multiple discrete aggregation states. *J. Mol. Biol.* **204**, 959–972 (1988).
16. DiCapua, E., Schnarr, M., Ruigrok, R. W., Lindner, P. & Timmins, P. A. Complexes of RecA protein in solution. A study by small angle neutron scattering. *J. Mol. Biol.* **214**, 557–570 (1990).
17. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).
18. Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
19. Yu, X., VanLoock, M. S., Yang, S., Reese, J. T. & Egelman, E. H. What is the structure of the RecA–DNA filament? *Curr. Protein Pept. Sci.* **5**, 73–79 (2004).
20. Levin-Zaidman, S. *et al.* Ordered intracellular RecA–DNA assemblies: a potential site of *in vivo* RecA-mediated activities. *Proc. Natl Acad. Sci. USA* **97**, 6791–6796 (2000).
21. Renzette, N. & Sandler, S. J. Requirements for ATP binding and hydrolysis in RecA function in *Escherichia coli*. *Mol. Microbiol.* **67**, 1347–1359 (2008).
22. Kidane, D. & Graumann, P. L. Dynamic formation of RecA filaments at DNA double strand break repair centers in live cells. *J. Cell Biol.* **170**, 357–366 (2005).
23. Gayathri, P. *et al.* A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* **338**, 1334–1337 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by an EMBO long-term fellowship to C.L. (LTF-535-2009); by a Wellcome Trust Programme Grant to D.J.S. (WT083469MA); and by a Wellcome Trust Strategic Award (091911) supporting advanced microscopy at Micron Oxford. We thank I. Dobbie for his support and advice on imaging. We thank D. Leach and S. Kowalczykowski for the gift of strains and plasmids, N. Dubarry for critical reading of the manuscript and M. Stouf for the gift of pMS11. We thank E. Birt and K. Prosser for preliminary experiments.

Author Contributions D.J.S. and C.L. conceived the project, interpreted the data and wrote the paper. C.L. designed and undertook experiments. G.B. developed analytical tools for microscopy data and provided advice on the manuscript. L.S. provided expertise on 3D-SIM and advice on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J.S. (david.sherratt@bioch.ox.ac.uk).

METHODS

Bacterial strains and growth. All strains are derivatives of *E. coli* K12 TB28 (MG1655; *ΔlacZYA*²⁴). Cells were grown at 30 °C in M9 media supplemented with 0.2% glucose. Ampicillin (100 μg ml⁻¹), kanamycin (50 μg ml⁻¹), chloramphenicol (20 μg ml⁻¹) and tetracycline (10 μg ml⁻¹) were added when required. Expression of I-SceI endonuclease from an arabinose-inducible promoter²⁵ was induced by the addition of 0.2% arabinose. Intracellular localization of chromosomal loci flanking the I-SceI^{CS} cut site was visualized by fluorescent wide-field microscopy using two distinct ParB/*parS* systems²⁶. Mutations were generally constructed by λ-Red recombination²⁷ and further introduced into the strain of interest by P1 transduction. When needed, the DNA region between the two *frt* sites (*cam/kan* resistance genes) was removed using FLP recombinase expressed from pCP20 (ref. 27).

Microscopy sample preparation. All microscopy experiments were performed on live cells growing in exponential phase ($A_{600\text{ nm}} \sim 0.1\text{--}0.2$). Cells were transferred from liquid culture to a slide mounted with 1% agarose in M9 glucose 0.2% medium and incubated during microscopy at the required temperature using an incubation chamber. 3D-SIM snapshots were performed at 24 °C with cells covered with precision cover glass, thickness no. 1.5H (170 μm ± 5 μm; Marienfeld Superior) using immersion oil with a refractive index of 1.512 to minimize spherical aberration. For time lapses after DSB induction, arabinose 0.2% was added to the liquid culture for 30 min before the cell were mounted on the slide. Nucleoids were visualized using 4 μg ml⁻¹ 4',6-diamidino-2-phenylindole (DAPI), and cell membranes were visualized using 1 μg ml⁻¹ FM4-64 (Life Technologies).

Wide-field microscopy, 3D-SIM and FRAP imaging. Conventional wide-field fluorescence microscopy was carried out on an Eclipse TE2000-U microscope (Nikon), equipped with a ×100/1.4 oil PlanApo objective and either a Cool-Snap HQ CCD or a QuantEM camera (Photometrics), and using Metamorph software for image acquisition. Super-resolution 3D-SIM imaging as well as conventional wide-field imaging was performed on a DeltaVision OMX V3 (Applied Precision/GE Healthcare) equipped with a Blaze SIM module, a ×60/1.42 oil UPlanApo objective (Olympus), 405 nm, 488 nm and 593 nm diode lasers and three sCMOS cameras (PCO). The fast-live mode enabled ultra-high-speed illumination and simultaneous or sequential acquisition of multiple-colour three-dimensional stacks of RecA-GFP/DAPI, RecA-GFP/FM4-64, RecA-GFP/mCherry or RecA-GFP alone. For each colour, the raw 3D-SIM stacks were composed of 225 images (512 × 512 pixels) consisting of 15 z-sections (125 nm z-distance, sample thickness of 1.750 μm), with 15 images per z-section with the striped illumination pattern^{28,29} rotated to the three angles (−60°, 0°, +60°) and shifted in five phase steps. Acquisition settings were as follows: RecA-GFP, 2–5 ms exposure with 488 nm laser (attenuated to 10% transmission); DAPI, 20 ms exposure with 405 nm laser (100% transmission); FM4-64, 30 ms exposure with 593 nm laser (100% transmission). Total acquisition times per stack were 1.8 s for RecA-GFP, and ~10 s for RecA-GFP/DAPI and RecA-GFP/FM4-64. Note that simultaneous imaging of RecA-GFP and DSe foci (ParB-mCherry) was performed in 3D-SIM mode for RecA-GFP and conventional wide-field mode for ParB-mCherry. The 3D-SIM raw data was computationally reconstructed with SoftWoRx 6.0 (Applied Precision) using a Wiener filter setting of 0.002 and channel specifically measured optical transfer functions to generate a super-resolution three-dimensional image stack with a lateral (x–y) resolution of 100–130 nm (wavelength dependent) and an axial (z) resolution of ~300 nm. In the reconstruction process the pixel size is halved from 80 nm to 40 nm and the pixel number doubled in order to meet the Nyquist sampling criterion. The number of pixels in the conventional images of DSe foci were doubled using Priism (Image Visualization Environment; <http://msg.ucsf.edu/IVE/>) in order to merge with 3D-SIM images. A constrained iterative three-dimensional image deconvolution was applied to conventional wide-field data in SoftWoRx 6.0. Images from the different colour channels were registered with alignment parameters obtained from calibration measurements with 0.2-μm-diameter TetraSpeck beads (Life Technologies) using the OMX Editor software (C. Weisiger and J. Sedat, unpublished). FRAP experiments were performed with an UltraVIEW VoX spinning disk confocal system with Photokinesis module (PerkinElmer) assembled on an IX8 microscope (Olympus) equipped with an C9100-13 EMCCD camera (Hamamatsu) and a ×100/1.4 oil PlanApo objective (Olympus). Photo-bleaching of a small diffraction-limited spot was carried out with a focused 488 nm laser beam (100 ms with AOTF set to 50% transmission). Fluorescence intensity measurements of the unbleached and bleached regions were performed using ImageJ (<http://rsbweb.nih.gov/ij/>). Values were normalized to those of initial pre-bleaching images.

Snapshot and time-lapse analysis. Snapshot analysis was performed with the MicrobeTracker suite³⁰ extended by custom MATLAB routines that we specifically developed to generate focus positioning dotplots, histograms of 0/1/2-focus cell fractions, two-colour cell-type counting, ISD, USe–DSe focus distance, focus position along the cell long x-axis (length) and short y-axis (width). Distributions of DSe foci along the cell diameter presented Fig. 1B were determined by subdividing

cells into five cell slices of equivalent areas as previously described³¹. Stoichiometry of RecA was determined by convolving the total fluorescence in *recA-GFP* cells and comparing it to that of *mukB-GFP* cells for which MukB stoichiometry is known³². Total intracellular fluorescence intensity distribution analysis was performed on images resulting from the average projection of 15 z-sections (125 nm z-distance) corresponding to a sample thickness of 1.750 μm, using MicrobeTracker. Our estimate of RecA stoichiometry is in the range of previous estimates^{33–36}. Fluorescent particle tracking during long time lapses (>30 s per frame) was performed using a semi-automated custom MATLAB routine as previously described³⁷. Mean square displacement (MSD) in x and y dimensions and directionality of movement were obtained by running the View5D plug-in (ImageJ) on short time lapses (300 s; 5 s per frame). The apparent two-dimensional diffusion coefficient (D_{app}) was calculated from the initial slope of MSD plotted against time (t) using $MSD_{XY} = 4(D_{app})t$. The directionality of movement reflects the fraction of the overall displacement (Δd_{i-}), compared to the total distance travelled by the focus ($\Delta d_1 + \Delta d_2 \dots + \Delta d_i$); that is, directionality = $(\Delta d_{i-})/(\Delta d_1 + \Delta d_2 \dots + \Delta d_i)$.

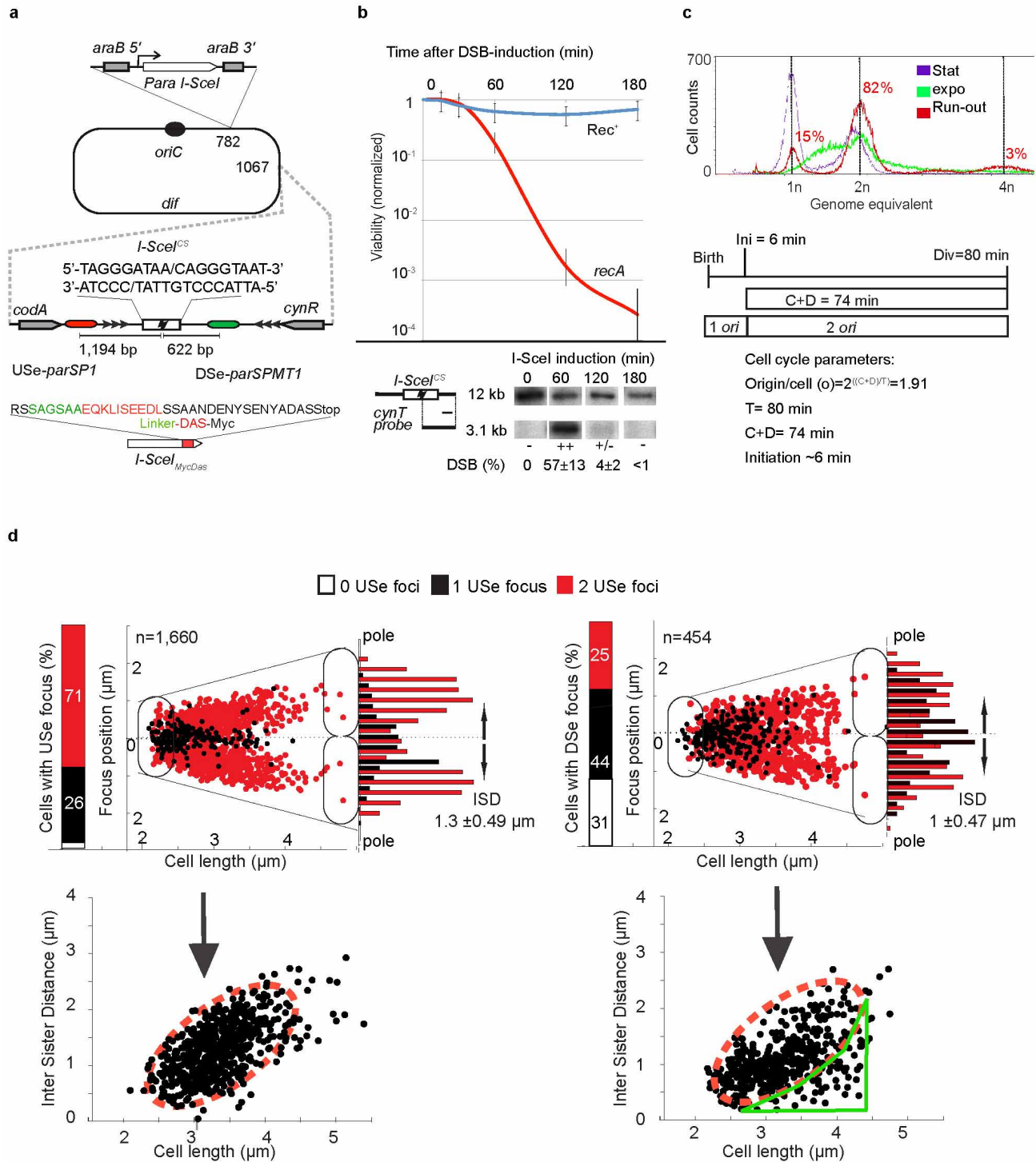
Immunocytochemistry. Cells were fixed in phosphate buffered saline with para-formaldehyde 5% (pH 7.2) and glutaraldehyde 0.06% and immobilized on poly-L-lysine slides (Poly-prep; Sigma-Aldrich), followed by treatment with lysozyme 1 μg ml⁻¹. Immunocytochemistry was performed using rabbit anti-RecA polyclonal antibodies (Abnova), which were revealed using anti-rabbit secondary antibodies conjugated to Alexa 594 (Life Technologies). Coverslips were mounted with Vectashield mounting medium (Vector Laboratories).

Flow cytometry. Cells were grown in M9 glucose at 30 °C and sampled in late stationary phase, exponential phase ($A_{600\text{ nm}} \sim 0.15$) and 3 h after incubation with cephalixin and rifampicin (run-out). Cell samples were prepared as described³⁸ except that Syto16 (Life Technologies) was used to stain DNA. Analysis was performed on FACScalibur flow cytometer (BD Biosciences). Data files were analysed using CellQuest (BD Biosciences) and Weasel (WEHI; <http://www.wehi.edu.au/>).

Introduction of DSBs by I-SceI. The Rec⁺ strain, in which DSBs were initially induced, corresponds to TB28 with both the I-SceI cut site and the gene encoding the I-SceI endonuclease inserted into *codA-cynR* and *araB* chromosome loci, respectively. The I-SceI^{CS} cassette carries *frt-cam-parS1-frt-3χ⁺-I-SceI^{CS}-parSMT1-3χ⁻*, where I-SceI^{CS} stands for the cut site (5'-TAGGGATAACAGGGTAAT-3'), + or – stands for the orientation of 5'-GCTGGTGG-3' χ-sites³⁹ and *frt* indicates the FLP site-specific recombination sites. *parS* sites were amplified from pGBKD3-*parSMT1* and pGBKD3-*parS1* plasmids⁴⁰. The cassette was integrated into the *codA-cynR* intergenic region of the MG1655 chromosome (356.6 kb) or into the *ydeO-ydeP* intergenic region (1,581.7 kb) by λ-Red recombination²⁷ and further transferred by P1 transduction into TB28. The wild-type allele of an I-SceI gene under the control of an arabinose-inducible promoter (carried by the pDL2655 plasmid described previously³⁹) was fused to 1× c-Myc tag followed by the DAS+4-degron tag in its C terminus⁴¹ amplified from pROD60. This plasmid carrying a 6-amino-acid linker, Myc tag, 2-amino-acid linker and the DAS+4 degron tag (SAGSAEEQKLISEED LSSAANDENYSENYADAS) was as described⁴². The I-SceI-degron construct was tested for its reduced efficiency to linearize an I-SceI^{CS}-carrying plasmid after induction with 0.2% arabinose *in vivo*. *P_{ara}*-I-SceI-*frt-kan* or *P_{ara}*-I-SceI-degron-*frt-kan* constructs were inserted in the *araB* gene (70 kb) of MG1655/pKD46 by the process of λ-Red recombination and further transferred into TB28 by P1 transduction. The *recA* gene with its *P_{recA}* promoter region were amplified from K12 genomic DNA and subsequently integrated downstream of the *fluB* gene on the MG1655 chromosome (171 kb on the chromosome) by λ-Red recombination and transferred into TB28 by P1 transduction (referred as *fluB::recA-cat*).

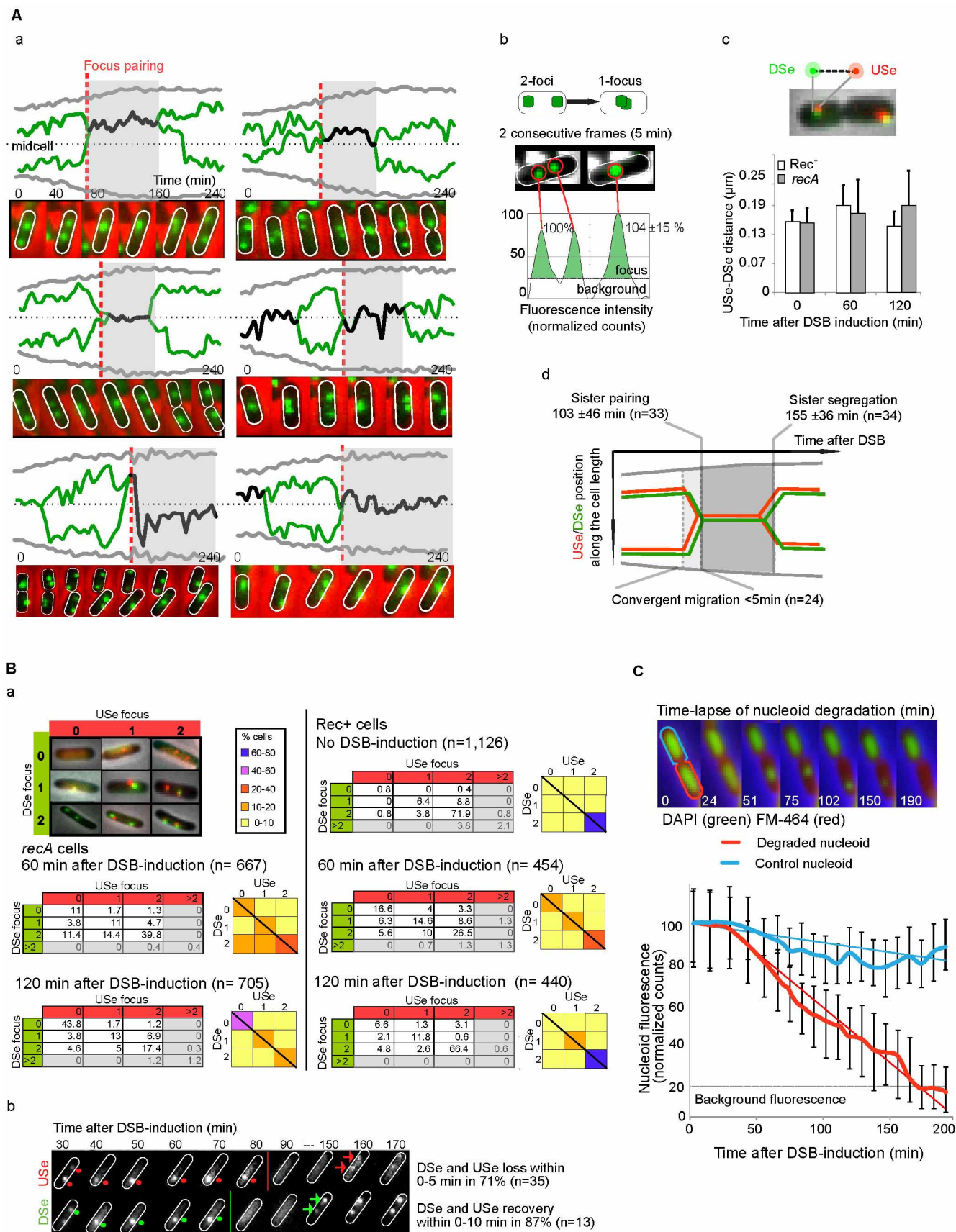
24. Bernhardt, T. G. & de Boer, P. A. Screening for synthetic lethal mutants in *Escherichia coli* and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity. *Mol. Microbiol.* **52**, 1255–1269 (2004).
25. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
26. Nielsen, H. J., Ottesen, J. R., Youngren, B., Austin, S. J. & Hansen, F. G. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol. Microbiol.* **62**, 331–338 (2006).
27. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
28. Gustafsson, M. G. *et al.* Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957–4970 (2008).
29. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).
30. Sliusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80**, 612–627 (2011).
31. Meile, J. C. *et al.* The terminal region of the *E. coli* chromosome localises at the periphery of the nucleoid. *BMC Microbiol.* **11**, 28 (2011).

32. Badrinarayanan, A., Reyes-Lamothe, R., Uphoff, S., Leake, M. C. & Sherratt, D. J. *In vivo* architecture and action of bacterial structural maintenance of chromosome proteins. *Science* **338**, 528–531 (2012).
33. Stohl, E. A. *et al.* *Escherichia coli* RecX inhibits RecA recombinase and coprotease activities *in vitro* and *in vivo*. *J. Biol. Chem.* **278**, 2278–2285 (2003).
34. Sommer, S., Boudsocq, F., Devoret, R. & Bailone, A. Specific RecA amino acid changes affect RecA–UmuD'C interaction. *Mol. Microbiol.* **28**, 281–291 (1998).
35. Sassanfar, M. & Roberts, J. Constitutive and UV-mediated activation of RecA protein: combined effects of recA441 and recF143 mutations and of addition of nucleosides and adenine. *J. Bacteriol.* **173**, 5869–5875 (1991).
36. Karu, A. E. & Belk, E. D. Induction of *E. coli* recA protein via recBC and alternate pathways: quantitation by enzyme-linked immunosorbent assay (ELISA). *Mol. Gen. Genet.* **185**, 275–282 (1982).
37. Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G. & Sherratt, D. J. Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc. Natl Acad. Sci. USA* **108**, E243–E250 (2011).
38. Lesterlin, C., Pages, C., Dubarry, N., Dasgupta, S. & Cornet, F. Asymmetry of chromosome replicores renders the DNA translocase activity of FtsK essential for cell division and cell shape maintenance in *Escherichia coli*. *PLoS Genet.* **4**, e1000288 (2008).
39. White, M. A., Eykelenboom, J. K., Lopez-Vernaza, M. A., Wilson, E. & Leach, D. R. Non-random segregation of sister chromosomes in *Escherichia coli*. *Nature* **455**, 1248–1250 (2008).
40. Espeli, O., Mercier, R. & Boccard, F. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.* **68**, 1418–1427 (2008).
41. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering controllable protein degradation. *Mol. Cell* **22**, 701–707 (2006).
42. Reyes-Lamothe, R., Sherratt, D. J. & Leake, M. C. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* **328**, 498–501 (2010).
43. Helmstetter, C. E. *et al.* in *Escherichia coli and Salmonella* Vol. 2 (ed. Neidhardt, F. C.) 1627–1639 (ASM, 1996).
44. Kuwada, N. J., Cheveralls, K. C., Traxler, B. & Wiggins, P. A. Mapping the driving forces of chromosome structure and segregation in *Escherichia coli*. *Nucleic Acids Res.* (2013).



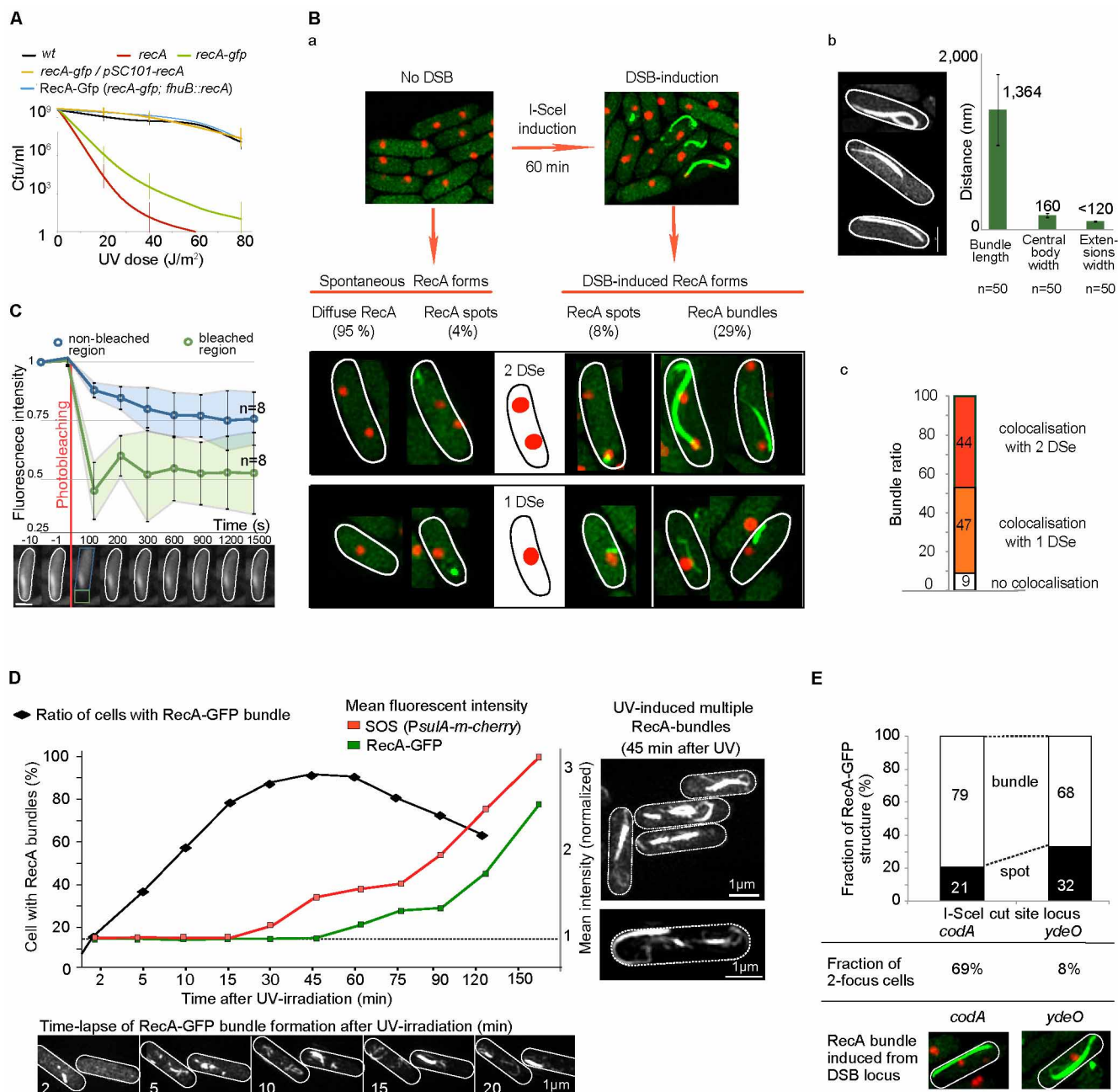
Extended Data Figure 1 | DSB induction. **a**, DSBs were introduced using the I-SceI endonuclease. The I-SceI cutting site (*I-SceI*^{CS}) is an 18 bp site (5'-TAGGGATAACAGGGTAAT-3') that was inserted 1,067 kb from *oriC* on the right replicore (*codA-cynR* intergenic region). *I-SceI*^{CS} is flanked by an upstream *parS-P1* site (USe; 1.2 kb upstream of the cut) and a downstream *parS-PMT1* site (DSe; 0.6 kb downstream of the cut) that binds different fluorescent ParB proteins. *In vivo* localization of USe and DSe was monitored in live cells by expressing the fluorescent fusions of the corresponding ParBP1- and ParBPMT1-binding proteins from pFHC2973 or pMS11. The *I-SceI* gene controlled by the arabinose-inducible promoter was expressed from the *araB* locus of the chromosome (782 kb from *oriC*), or alternatively expressed from a plasmid when a higher level of cutting was required. Two alleles of the *I-SceI* gene were used, namely the *wt* allele and an unstable allele carrying a degron tag (*I-SceI*_{MycDas}). The construction of the system is detailed in Methods. **b**, Cell viability of *Rec*⁺ and *recA* cells was assayed by plating efficiency after induction of *I-SceI*_{MycDas}. Colony-forming units (c.f.u. ml⁻¹) were counted 0, 30, 60, 120 and 180 min after addition of 0.2% arabinose in a steady-state population (OD_{600 nm} = 0.2). Values at *t* = 0 were normalized to 1. Bottom, Southern blot shows the efficiency of DSB induction for the same time points, and the level of DSBs measured from three independent experiments. A *cynT* probe (500 bp) was used to visualize the substrate: a 12 kb *SacII* fragment containing the *I-SceI* cut site, or the product of restriction by I-SceI (downstream 3.1 kb fragment).

Conditions were used in which 60 min after I-SceI was induced, most cells with segregated sisters received no more than a single DSB, consistent with the overall level of cutting observed biochemically (~57%) and the dependence on RecA for survival. **c**, Flow cytometry showing the DNA content distribution of TB28 strains in the stationary phase, during exponential growth (M9 glucose 30 °C) and after rifampicin/cephalexin run-out. Subpopulations harbouring 1 or 2 chromosome equivalents after run-out indicate the number of replication origins per cell (1.91), which was used to calculate the *C* (DNA synthesis period) + *D* (time between completion of DNA replication and cell division) period (74 min) using Helmstetter's model (origin per cell = 2^{(C+D)/t}) (ref. 43). A summary of the cell cycle parameters is shown. All experiments were performed in identical growth conditions in which cells harbour only one replicating chromosome, with initiation of replication occurring ~6 min after birth in the *T* = 80 min cell cycle. **d**, Snapshot analysis of USe localization before (left) and after (right) 60 min of DSB induction showed an identical pattern compared with DSe (Fig. 1A). From left to right: ratio of cells with 0, 1 or 2 foci; dot plots of USe foci positions as a function of cell length; histograms of USe position along the cell long axis and the ISD (in µm) are shown. Bottom, a plot of ISD against cell length shows that DSB induction triggers the global decrease of ISD values and generates a population of long cells with reduced ISD (framed in green), revealing the pairing of sister loci.



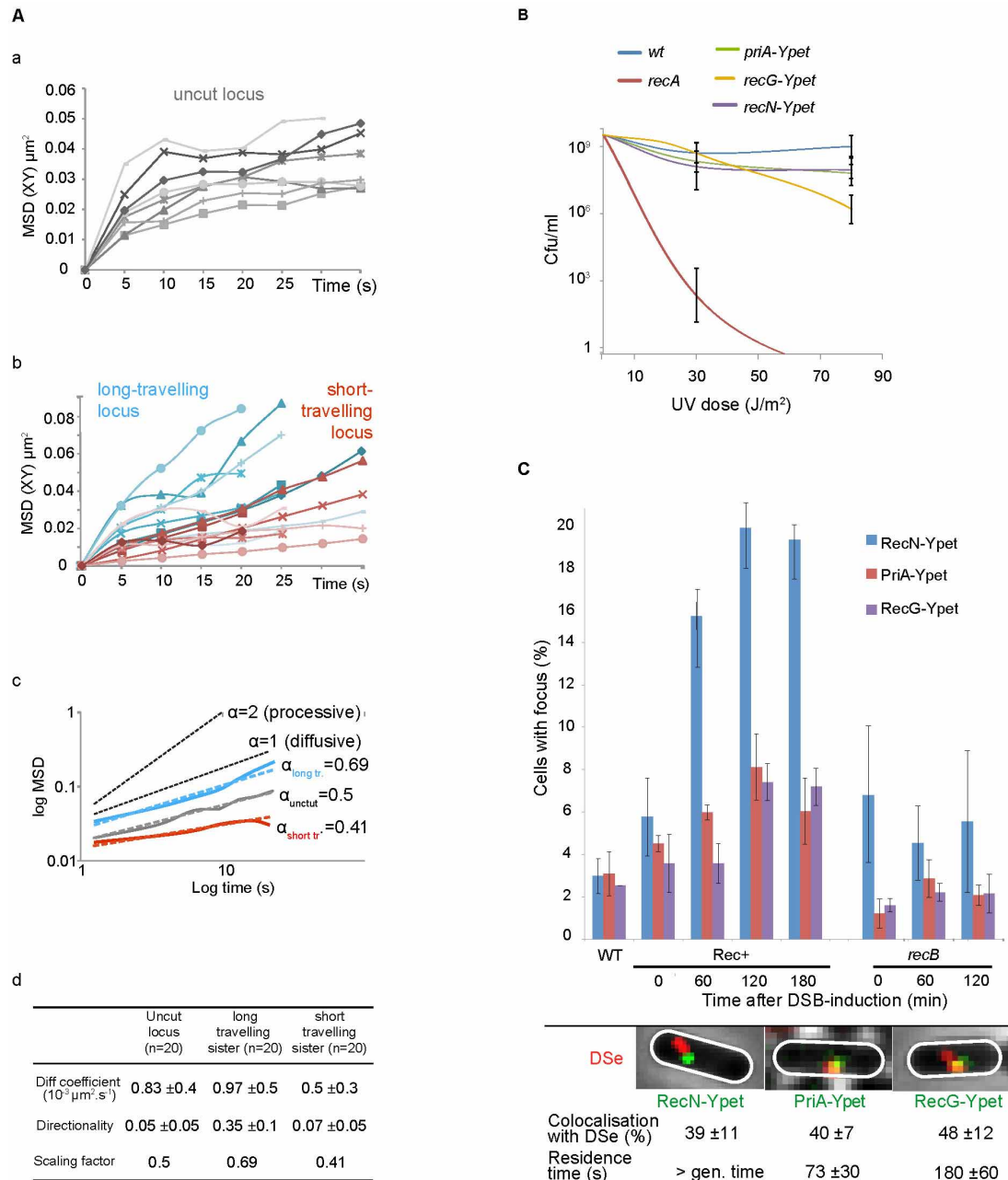
Extended Data Figure 2 | Consequences of DSB induction. **A**, Analysis of DSB-induced DSe sister pairing in Rec^+ cells. **a**, Time-lapse particle tracking of DSe foci (5 min per frame) with indicative images (every 20 min) showing pairing between segregated sister loci (DSB was induced 30 min before the beginning of time-lapse acquisition). The red dashed line indicates the sister-pairing event and the grey area indicates the pairing period. **b**, DSe focus fluorescence intensity quantification performed on two consecutive frames—that is, before and after pairing—shows that the one focus observed after pairing retained $104 \pm 15\%$ of the added fluorescence of the previous sister focus (normalized at 100%), consistent with colocalization of the two localization arrays. **c**, Distances between DSe and USe focus centres remained unchanged upon DSB induction in Rec^+ and *recA* cells, indicating that both DNA ends remain coincident throughout cutting, pairing and repair. **d**, Summary diagram of sister pairing that occurred 103 ± 46 min after DSB induction and lasted for 50 ± 34 min (grey area). Once initiated, pairing occurred rapidly (<5 min) and engaged both DSB ends. **B**, **a**, DSB induction generated a variety of cell types harbouring 0, 1 or 2 DSe foci with 0, 1 or 2 USe due to loss of the DSe and/or USe foci. One image for each of the nine cell types is shown with their quantification 0, 60 and 120 min after DSB induction in Rec^+ and *recA* strains.

These data are displayed using a square density representation. This was obtained with a custom MATLAB code that uses the data generated by the spotfinderM function of MicrobeTracker software. Only cells $>3.3 \mu\text{m}$ are considered in order to focus the analysis on the fraction of cells in which the I-SceI cut site region (*codA-cynR*) has already segregated. **b**, Two-colour time-lapse analysis after DSB induction showed that DSe and USe foci belonging to the same DSB disappeared within 0–5 min from one another in 71% of cases (green and red lines), and recovered within 0–10 min from one another in 87% of cases (green and red arrows), consistent with bidirectional processing of the two ends of the DSB by RecBCD (see Supplementary Discussion). **C**, Time-lapse analysis of RecBCD-mediated complete degradation of the nucleoid after DSB induction in *recA* cells. One example is shown with nucleoid DNA stained with DAPI (green), cell membranes stained with FM4-64 (red). The decrease in fluorescence intensity of nucleoids undergoing degradation (red curve) was measured and reflects the rate of DNA degradation *in vivo* (Supplementary Video 1). Decrease in fluorescence of non-degraded nucleoids is accounted for by photobleaching of the DAPI (blue curve). The overall rate of DNA degradation was 23.2 kb min^{-1} (386 bp s^{-1}); that is, 193 bp s^{-1} per RecBCD complex ($n = 38$ cells analysed).



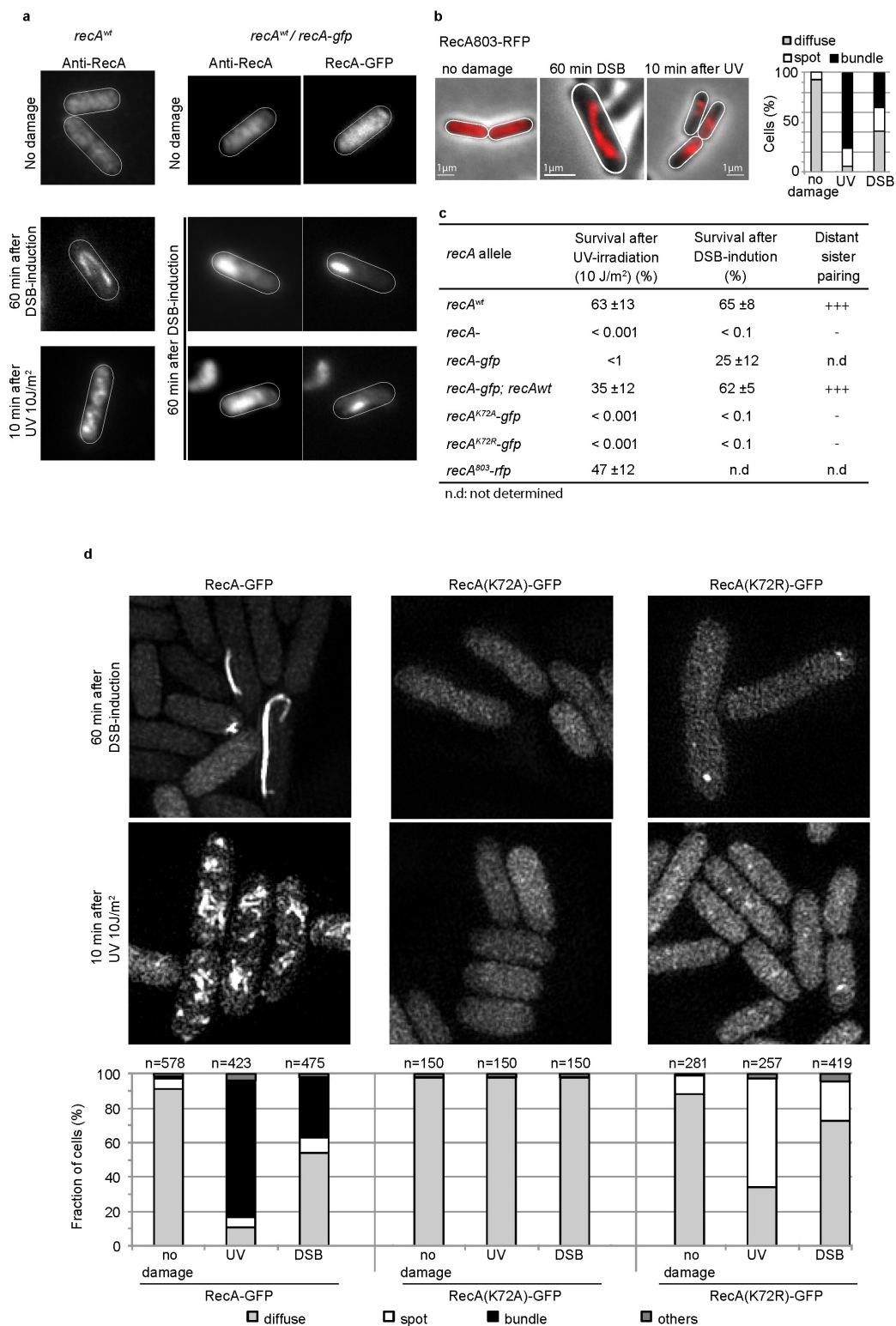
Extended Data Figure 3 | RecA bundle characterization. **A**, Survival after ultraviolet (UV) irradiation shows complementation of *recA-gfp* allele sensitivity (SS0385 strain) by expression of a wild-type *recA* gene in the *fhuB* ectopic locus of the chromosome or from a plasmid (pSC101-*recA*). RecA-GFP strain (TB28, *recA-gfp*, *fhuB::recA*) showed a level of UV resistance similar to the wild-type (*wt*) strain. **B**, **a**, Snapshot analysis of RecA-GFP structures induced by DSBs. Each image was obtained by projection of a three-dimensional stack (15 z-sections of 125 nm z-distance) of 3D-SIM images for RecA-GFP (green) merged with wide-field images for DSe (red). Before DSB induction, RecA is diffuse in most cells (95%) and spontaneous spots are distant from DSe foci. Sixty minutes after DSB induction, RecA spots form in close proximity to a DSe focus and RecA bundles are observed in 29% of total cells, regardless of whether cells have one or two DSe foci. **b**, Bundle length (nm), and the thickness of the large central body and the thinner extensions is shown in the histogram on the left (nm). The width of the thinner extensions (≤ 100 nm) corresponds to the lateral resolution limit of 3D-SIM (see Supplementary Videos 2–7). **c**, In cells with two DSe foci and a RecA bundle, 47% of RecA-GFP bundles were associated with one DSe focus and 44% were associated with both DSe foci. **C**, FRAP of RecA bundles. Fluorescence intensities of bleached (green

curve) and non-bleached regions (blue curve) of RecA bundles were monitored up to 1,500 s after bleaching. No fluorescence recovery was observed throughout this period, as shown by the indicative time-lapse images below. The green and blue boxes indicate the bleached and non-bleached regions of the bundle, respectively (time in s). Initial fluorescence was normalized to 1 (Supplementary Video 3). **D**, RecA bundle formation after UV irradiation. The ratio of cells with RecA bundle(s) after irradiation with UV light (50 J m^{-2}) was measured, as was the total fluorescence intensity of RecA-GFP. The level of the SOS response, induced in response to DNA damage, was reported by the expression of the *m-cherry* gene controlled by the *PsuIA* SOS-inducible promoter. Time-lapse images of RecA-GFP after UV irradiation confirms rapid bundle formation after UV treatment and snapshot images show multiple bundles present in most cells 45 min after UV irradiation (Supplementary Video 8). **E**, Relative frequencies of RecA bundles and RecA spots 60 min after DSB induction at *codA* (356.6 kb) or *ydeO* (1,581.7 kb) cut sites. *ydeO* and *codA* loci have different replication-segregation timings as manifested by the fractions of two-focus cells (8% and 69%, respectively). Images below show the morphology of RecA bundles observed in those two cell populations (see Supplementary Discussion).



Extended Data Figure 4 | Movement of loci after DSB induction and recruitment of recombination proteins. **A, a**, Mobility analysis of uncut loci. **b**, Comparative mobilities of long-travelling and the short-travelling sister foci during DSB-induced sister pairing. **c**, Mean square displacement values (MSDs), and MSD scaling parameters (α); $\alpha = 0.5$ for the uncut locus compared with 0.69 for the long-travelling sister and 0.41 for the short-travelling one. This compares with $\alpha = 1$ for diffusive movement and $\alpha = 2$ for processive movement⁴⁴. **d**, Summary showing the diffusion coefficients D_{app} , the directionality of movement and the scaling factor values for uncut loci, and long- and short-travelling loci. **B**, Survival after UV irradiation of cells expressing RecN-YPet, PriA-YPet and RecG-YPet fluorescent proteins

expressed from their endogenous promoters in their normal chromosomal position. All four fusion proteins were recombination proficient, as demonstrated by resistance to UV irradiation. **C**, Histogram of the fraction of cells harbouring RecN-YPet, RecG-YPet or PriA-YPet fusions showing DSB-induced focus formation in *Rec*⁺, but not in *recB* cells. Averages and standard deviations are calculated from snapshot images of three independent experiments. Bottom, the percentage of RecN-YPet, PriA-YPet and RecG-YPet foci that colocalized ($<0.5 \mu\text{m}$ apart) with a DSe focus is shown with indicative images. Residence times (s) of these foci, estimated by time-lapse analysis, are indicated (see Supplementary Discussion).



Extended Data Figure 5 | RecA bundles are not artefacts and require that RecA can bind ATP. **a**, Immunocytochemistry using anti-RecA antibodies and Alexa-594-conjugated secondary antibodies. Left, bundles formed by wild-type RecA 60 min after DSB induction (in >25% of analysable cells) or 10 min after UV induction (in >80% of analysable cells). In the *recA^{wt}/recA-gfp* diploid strain, bundles were visualized both by immunocytochemistry and by direct observation of RecA-GFP. **b**, Visualization of RecA803-RFP upon DNA damage induction. Merge of phase contrast and RFP images acquired with conventional fluorescent microscopy are shown. RecA803-RFP appeared diffuse in the cell before damage induction, and showed bundle formation

similar to that observed with the RecA-GFP fusion after DSB induction (same conditions as in Figs 1 and 2) or UV irradiation (10 J m⁻²). Histogram presents the fractions of cells with RecA803-RFP bundles in corresponding conditions. **c**, Survival of cells carrying different *recA* alleles after DSB induction and UV irradiation (10 J m⁻²), estimated by a plating assay. **d**, 3D-SIM images of RecA-GFP, RecA(K72A)-GFP and RecA(K72R)-GFP fusions after DSB induction (same conditions as in Figs 1 and 2) or UV induction (10 J m⁻²). Histograms present the fraction of cells with diffuse fluorescence, cells with RecA spots or RecA bundles ('others' corresponds to cells with saturated signal or aberrant morphologies).

CORRIGENDUM

doi:10.1038/nature13003

Corrigendum: IGFBP-4 is an inhibitor of canonical Wnt signalling required for cardiogenesis

Weidong Zhu, Ichiro Shiojima, Yuzuru Ito, Zhi Li, Hiroyuki Ikeda, Masashi Yoshida, Atsuhiko T. Naito, Jun-ichiro Nishi, Hiroo Ueno, Akihiro Umezawa, Tohru Minamino, Toshio Nagai, Akira Kikuchi, Makoto Asashima & Issei Komuro

Nature **454**, 345–349 (2008); doi:10.1038/nature07027

It has recently been brought to our attention that some of the lanes in the reverse-transcriptase-mediated polymerase chain reaction (RT-PCR) analyses of this Letter appear to be duplicated. Specifically, two β -actin bands in Fig. 1c, lanes 1, 4 and 6 of the α MHC bands and lanes 1 and 6 of the GATA4 bands in Fig. 3e, lane 1 of the Nkx2.5 band and lane 3 of the GATA4 band in Supplementary Fig. 3e, and lane 1 of the Nkx2.5 band and lane 1 of the β -actin band in Supplementary Fig. 3f might be duplicated. Lane 2 of the GATA4 band in Supplementary Fig. 3c might be 'pasted in'. Because the original raw data could not be located, we were unable to verify definitively whether the data in the original figures are correct or whether some of the lanes were indeed inadvertently duplicated. We therefore repeated the RT-PCR experiments shown in these figures. There is one difference in the experimental procedures between the original work and the repeated experiments. In the original work, knockdown of insulin-like growth-factor-binding protein 4 (IGFBP-4) in P19CL6 and embryonic stem (ES) cells was achieved by establishing stable cell lines that constitutively express IGFBP-4 small interfering RNAs (BP4-1 and BP4-2 in Fig. 3e and Supplementary Fig. 3e). In the repeated experiments, knockdown of IGFBP-4 was done by transient transfection of IGFBP-4 siRNAs. Figure 1 of this Corrigendum shows the corrected right-hand panels of the original Figs 1c and 3e. The Supplementary Information of this Corrigendum shows the corrected Supplementary Fig. 3c, e and f of the original Letter (in which the RT-PCR panels represent new and independent experiments); it also contains the raw data for all the corrected figures. Repeated experiments show that the results are reproducible and essentially the same as those of the original Letter. We thank Seitaro Nomura from the University of Tokyo, who generated data for the corrigendum. We apologize for any confusion these possible errors may have caused. The corrections described here do not alter the overall conclusion that IGFBP-4 promotes cardiogenesis by inhibiting canonical Wnt signalling.

Supplementary Information is available in the online version of this Corrigendum.

Correspondence should be addressed to I.K. (komuro_tky@umin.ac.jp).

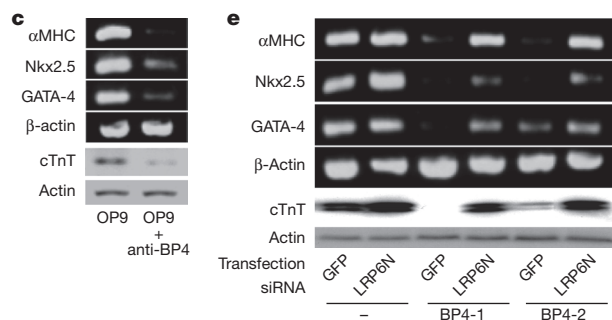


Figure 1 | This figure shows the corrected right-hand panels of Fig. 1c and Fig. 3e of the original Letter. In both figures, the RT-PCR panels represent new and independent experiments.

CORRIGENDUM

doi:10.1038/nature13042

Corrigendum: The architecture of *Tetrahymena* telomerase holoenzyme

Jiansen Jiang, Edward J. Miracco, Kyungah Hong, Barbara Eckert, Henry Chan, Darian D. Cash, Bosun Min, Z. Hong Zhou, Kathleen Collins & Juli Feigon

Nature **496**, 187–192 (2013); doi:10.1038/nature12062

Owing to an editorial oversight, accession numbers for the negative-stain electron microscopy maps for this Article were not obtained. The electron microscopy maps have been deposited in the Electron Microscopy Data Bank (<http://emdatbank.org>) for all of the structures shown in the main text and in the Supplementary figures. The accession codes are 5804 (the stable major conformation of Teb1-F telomerase analysed in detail) and 5807–5824 (from minor conformations of Teb1-F telomerase, TERT-F telomerase, and Fab-labelled TERT-F, Teb1-F and F-p50 telomerase).

ERRATUM

doi:10.1038/nature13041

Erratum: Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells

Yukihiro Furusawa, Yuuki Obata, Shinji Fukuda,
Takaho A. Endo, Gaku Nakato, Daisuke Takahashi,
Yumiko Nakanishi, Chikako Uetake, Keiko Kato, Tamotsu Kato,
Masumi Takahashi, Noriko N. Fukuda, Shinnosuke Murakami,
Eiji Miyauchi, Shingo Hino, Koji Atarashi, Satoshi Onawa,
Yumiko Fujimura, Trevor Lockett, Julie M. Clarke,
David L. Topping, Masaru Tomita, Shohei Hori, Osamu Ohara,
Tatsuya Morita, Haruhiko Koseki, Jun Kikuchi, Kenya Honda,
Koji Hase & Hiroshi Ohno

Nature **504**, 446–450 (2013); doi:10.1038/nature12721

In Fig. 1d of this Letter, the vertical label of the upper right panel should be “Neuropilin-1⁺ Foxp3⁺ (%)” instead of “Neuropilin-1⁻/Foxp3⁺ (%)”; this error has been corrected in the online versions of the paper.

CAREERS

COLUMN Junior researchers can learn soft skills by forming their own groups **p.257**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com



WEB DESIGN

Webcraft 101

An eye-pleasing website can boost the appeal of a laboratory, and creating one has never been easier.

BY ROBERTA KWOK

Emilio Bruna admits that his first laboratory website was painfully ugly. Bruna, a plant ecologist at the University of Florida in Gainesville, put together the site in 2002 using Microsoft FrontPage software. The background colour was a homely tan, the text was unevenly indented and a news ticker advertising for research-assistant applicants crawled along the bottom of the page. Bruna calls the look “horrible”. He later upgraded to a university website template, but he disliked the generic corporate look.

He was never really satisfied with the design until he started using the free website-building tool WordPress last year, which allows users to

choose from thousands of designs, edit from any computer and easily customize their sites with ready-made modules such as an automatically updated Twitter feed. Bruna chose a clean, attractive template and added elements such as a slideshow of current projects and blog posts. Although he acknowledges that his current site (<http://brunalab.org>) is still not the pinnacle of sophisticated design, he calls it a major improvement on the previous version. The new site has attracted admiring comments from colleagues, and Bruna believes that it offers a clearer, more appealing presentation of his research that provides other scientists with easy access to his papers and data.

Many scientists consider creating and maintaining an aesthetically pleasing, up-to-date

website a low priority. But a compelling lab homepage can help to attract students and funding. Some students browse potential advisers’ websites to help them decide whether the lab culture is a good fit, and scientists sometimes check other researchers’ sites to keep up to date on a lab’s work or to decide whether they want to collaborate. Thoughtful design also makes it easier to read and process information.

“It’s so important to have an appealing current webpage because this is your professional presence online,” says Liz Neeley, who is based in Seattle, Washington, and is assistant director of science outreach at the science communication organization COMPASS. “It’s the face that you’re showing to the broader world.”

Although social-media sites have gained importance as communication tools, a website remains a convenient, centralized source of information about projects, publications, research products and teaching — as well as an opportunity to showcase what makes the lab unique. By carefully organizing their content, considering their target audience and following some simple design guidelines (see ‘Digital delight’), scientists can craft a site that presents their accomplishments and scientific approach. And with many user-friendly tools now available, building a new website does not have to be expensive or onerous. “There’s no excuse for having an ugly webpage anymore,” says Bruna.

CRAFTING THE CONTENT

At a minimum, says Neeley, every lab website should have a concise one- to two-paragraph research summary, a list of lab members with titles and biographies, and a head shot of the principal investigator. Lab websites also often list research projects, publications, data sets, software, job openings, collaborators and contact information.

For more in-depth content, Neeley suggests a layered strategy. An overview summary on the first page could emphasize key projects and other issues that the principal investigator considers to be important, such as public outreach or open science. A research page could provide project descriptions in layman-friendly language, with links to papers or more technical information. This allows interested readers to get a range of information without diving into details right away, says Neeley. Be selective about which projects are featured on the front page; if too many are included, the less important ones “might be taking attention away from what really is relevant,” says Gaby Hernández, a graphic-design researcher and educator ►

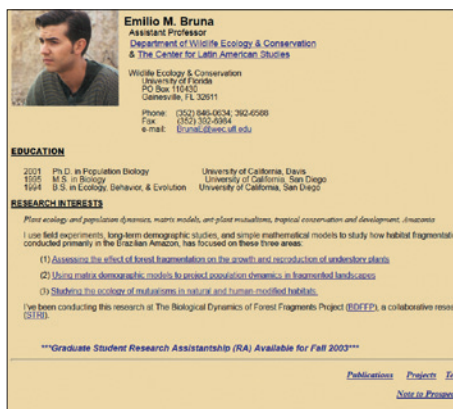
RATCH/Shutterstock; SCREENSHOT: R. DUNN/N. MCCOY

► at the University of Houston-Downtown in Texas, who has worked with scientists on communication and data visualization.

When crafting the site, consider who it is targeting. Astronomer David Kipping used to have a somewhat crowded, blog-like website aimed at a popular audience. But when Kipping, who is currently at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, entered the job market last year, he wanted to tailor the site towards busy faculty members reviewing his applications. So with guidance from design-savvy family members, Kipping reworked the site (www.davidkipping.co.uk) to look more like an online CV, with easy-to-find pages for papers and project descriptions, and newly added information about his teaching record. His list of publications included links to the papers, as well as associated animations and 'cheat sheets' of equations that convey the main points of the studies (because many researchers might not read the entire paper). The goal was to "help the reviewer to very quickly find the relevant information," says Kipping. And, he notes, an appealing website "can also show that you are capable of putting together a nice presentation and outreach tool".

Details or illustrative flourishes can entice prospective students. Bruna added the slideshow of projects to his website, including images of his team's work in Brazil, partly to attract students who are interested in tropical ecosystems. The 'lab members' section mentions the activities of lab alumni and includes e-mail addresses for current students, so prospective students can contact them for input about the lab. To reduce time spent responding to e-mail enquiries, Bruna included instructions on what information prospective students should send; most applicants follow those guidelines, he says.

But sites can use more than research summaries or publication lists to attract prospective students: they can articulate a lab's philosophy and mission, helping to set the tone for the lab's culture. Neuroscientist Kay Tye signed a contract to become an assistant



Emilio Bruna calls his first lab website (left) "horrible"; the updated version (right) features a slideshow.

professor at the Massachusetts Institute of Technology in Cambridge in 2011, and soon after, she set out to recruit the right people for her new laboratory. She put up a lab website (www.tyelab.org) that included a philosophy statement describing her expectations for lab members; she emphasized the importance of having a positive attitude, communicating honestly, supporting other lab members and taking care of oneself (for example, by making time to socialize and exercise). Many applicants have cited the philosophy statement as a reason they wanted to work with her, says Tye. And stating expectations up front deters applicants who might not be a good fit. "People who don't identify with this philosophy won't apply, and then we won't waste our time," she says.

Some researchers even add website content to attract funding. Bruna applies primarily to the US National Science Foundation in Arlington, Virginia, for grants, and the agency has emphasized the importance of sharing data. So Bruna's site includes a statement supporting such policies and clear links to his data sets. "If you're an applicant for a grant, this is your opportunity to sell yourself," he says. If funders are interested in social impact, images of the scientist working with, for example, farmers affected by the research could be helpful, says Hernández. A website can also assist with crowd-funding.



In 2012, Ethan Perlstein, an independent evolutionary pharmacologist and founder of the research company Perlstein Lab in San Francisco, California, published several blog posts on his site (now at www.ethanperlstein.com) to promote a campaign to fund his amphetamine research. The campaign ultimately raised more than US\$25,000 (see *Nature* **490**, 133; 2012).

ATTENTION TO DESIGN

To get started, look at other people's websites and identify appealing elements, says Kipping. Researchers might find inspiration at the Tumblr page Great Lab Websites (<http://go.nature.com/iwn9kn>), which was started this year by biologist Michael Eisen at the University of California, Berkeley, or at the web-design gallery site Inspire (www.siteinspire.com).

Consider the personality or values that the site should express. Perlstein is an open-science proponent and heavy social-media user, so his front page features a stream of blog posts, videos and automatically updated content from his accounts on data sharing and social media sites. Other researchers might prefer a more conventional look. "I don't think there's one size fits all," he says.

Choose images that are specific to the research. For example, ecologists often use pictures of forests, but a close-up of the plant species being studied would be more unique, says Hernández. Ensure that images are in focus and have enough contrast to clearly show the subject.

Simple fonts such as Helvetica are easier to read than ornate fonts, and dark grey text generally looks better than black, which can appear too heavy, adds Hernández. A light background is easier on the eyes, and the colour of links should be clearly distinguishable from non-linked text. When choosing the link colour, stick with the site's existing colour palette; for example, a website with blue border lines could also have blue links.

But scientists don't need design expertise to create an appealing website — many 'ready-made' website templates are available. WordPress is one of the most popular sources of free design templates; Kipping used a template

DIGITAL DELIGHT

How to create a clear and attractive website

- Avoid crowding the home page with too much information, which can distract the reader from the most important content.
- Structure content in layers. Give a brief overview statement, then allow the reader to click to other pages for more details.
- Minimize the number of clicks needed to reach important information. Include clear links to source material, such as paper PDFs, data sets and software.
- Consider the target audience for the site, such as prospective students or hiring

- committees, then add or emphasize elements that are relevant to that group.
- Include a clear head shot of the principal investigator.
- Select images specific to the research, such as the plant species being studied, rather than a generic forest photo.
- Use simple fonts. Dark grey text on a light background is easy to read.
- Ensure that the link colour is clearly distinguishable from non-linked text. Stick to a consistent colour palette. **R.K.**

from the Mac program RapidWeaver; and the online service Squarespace offers slick designs and web hosting starting at \$8 per month. Whatever template they choose, researchers should ensure that the design has been optimized for mobile devices with, for example, alternative layouts that are easier to view on a smaller screen.

Costs for a customized site can vary depending on the requirements. Neil McCoy, a freelance web designer in Raleigh, North Carolina, who specializes in science communication, charges \$60 per hour and says that an academic site usually costs between \$1,000 and \$3,000; sites at the higher end of the price range might involve custom graphics or a large amount of content, such as many pages or photo galleries. Perlstein paid a company about \$8,000 for his website, which has advanced features such as the automatic import of content from alternative metrics sites. The design process, complete with mock-ups and a test site for Perlstein's review, took about three months. However, for scientists on a budget, graphic-design students might be willing to build a simple site relatively cheaply.

Improving a site doesn't have to take a lot of time. Bruna spent a weekend setting up his WordPress site and now spends about an hour per week updating it with news. But updates every couple of months,

"Like it or not, we are in the business of marketing and promoting our work."

when papers are published or accepted, will probably suffice for most scientists. Researchers could reduce maintenance time by displaying only representative publications rather than a comprehensive list. If overhauling the entire site sounds overwhelming, focus on the improvement that would yield the biggest gain, such as writing an overview statement or updating the publications list, says Neeley. Finishing a simple site is better than leaving a complicated one under construction; a website should not look like an abandoned project, says Attila Csordas, a bioinformatician at the European Molecular Biology Laboratory's European Bioinformatics Institute in Hinxton, UK, who helped to judge a 2007 lab-website competition hosted by *The Scientist*.

The rise of social media means that websites bear less of the burden of representing the lab than they did several years ago, says Csordas. Even so, the lab website remains a key element of a scientist's public profile. "Like it or not, we are in the business of marketing and promoting our work," says Bruna. "Those things do matter." ■

Roberta Kwok is a freelance science writer in Seattle, Washington.

COLUMN

Self-taught soft skills

Junior researchers can learn career-boosting skills by forming their own groups, says **Alexandra Lucs**.



Junior researchers are becoming increasingly aware that in addition to honing their scientific know-how, they need to develop their 'soft skills', such as the art of communicating, managing or collaborating. For most, opportunities for formal training in soft skills are limited. Independent courses are starting to emerge, such as the 'Leadership in BioScience' sessions at Cold Spring Harbor Laboratory in New York, but not everyone can afford the requisite time or money.

At the Feinstein Institute for Medical Research in Manhasset, New York, several young investigators and I are trying a different approach. We have found an easy and inexpensive way to foster our people skills by taking on the training ourselves. By forming a small group, we have learned soft skills from each other, as well as from resources found online.

We decided to include no more than one person from any given laboratory to avoid intralab politics and to ensure that members feel free to speak openly and confidentially about any personnel issues that might arise. This is particularly important when discussing, for example, what makes a good manager, the ways in which researchers influence others around them or the tools needed to manage difficult conversations.

Some were initially sceptical; they simply hadn't considered the importance of soft skills in a laboratory. But after discussing the need to better negotiate our future positions, hire a strong talent pool and mentor or manage current personnel, the doubters quickly understood the potential benefits.

Our ten-person group has met monthly since April 2013, with each meeting lasting 1.5 hours. We begin the sessions with brief updates on each

member's career progress and challenges, such as their grant applications, invitations to talk and problems in hiring capable personnel.

An online tool called Lean In (<http://leanin.org>) has greatly facilitated our efforts. Although intended for women, the website's informative videos feature topics that are relevant for both genders. They address such issues as 'How do team dynamics affect scientific contributions?', 'How can we exert influence in our collaborations?', 'Who at our institutions can help us learn how to become group leaders?' and 'How can we negotiate for new positions?' We watch the videos on our own and discuss them as a group.

One topic that we discussed was when to 'walk away': that is, how to recognize when a position or job offer does not fulfil your aims. The goal of one of our members was to become a professor, but she identified some additional goals, such as independence, upward mobility and the ability to pursue science research. She received a job offer that seemed to satisfy these requirements, but she didn't jump on it — not yet. She knew she could walk away if need be. At her next meeting she firmly stated that she needed her academic appointment within a certain time frame. There were no threats, only calm determination, and she got what she wanted.

Another group member recognized that her often-docile demeanour led to her contributions being under-appreciated. It inspired her to become a better advocate for her career, and she quickly got two job offers. Before joining the group, she would simply have accepted the first offer she received, but instead she negotiated with specific salary and time commitment requirements in mind. Both employers agreed.

An unanticipated benefit of this group is that we have formed a strong network. We talk to each other more outside meetings and have formed interdisciplinary collaborations. In attempting to learn soft skills together, we have also enhanced each other's science.

In the past few months, new people have joined our group to replace those who have moved on. Some may enter the group with doubts about its merit. But soon they realize just how important these soft skills are — and just how much we can teach ourselves. ■

Alexandra Lucs is a postdoctoral fellow at the Feinstein Institute for Medical Research in Manhasset, New York.

BRIGID SURGEON

HOWARD LOVES POLLY

In search of something real.

BY MARTIN HAYES

Howard and Polly held hands as they made their way through the pines and firs to the cliff edge where the strange lights had been seen for three weeks running. The crazy-looking science guy on the news had said that similar lights had been observed in more than 11,000 locations worldwide; even the Halley VI research station in Antarctica had reported sightings over the Weddell Sea.

The city's edgelands were eerily quiet; most people had gravitated towards the built-up centres where the gathered multitudes offered a comforting, if false, glimmer of security. Others just stayed at home with loved ones to wait and hope and maybe pray that what was coming wasn't really coming.

As they trudged up the forest trail, Howard couldn't keep the thought out of his mind that he never again wanted to let go of this sweaty little hand. They were both out of breath when the track levelled off and gave way to a wide stretch of grass and then bare rock that led to the cliff's edge.

They stopped a few feet from the precipice. Howard kicked a small stone into the darkness but didn't hear it land. He could see across the ravine to the cliff on the opposite side. There were people there too, stick-like in the twilight distance. Maybe ten in all.

For weeks the talking heads on TV had been telling the world that no one knew what the cause of the increasingly strange phenomenon was, or what the outcome was likely to be, but deep down almost everyone had a bad, queasy feeling in their guts. And then last night they'd told it all, to the whole world, because what was the point in keeping it secret any more? The sightings had been increasing exponentially, and some lank-haired numbers-geek had worked out that if something was going to happen, it was going to happen tonight. And that something was going to be bad.

Howard took a hip flask from his jacket and unscrewed the cap. He offered Polly a drink and she made a funny face.

➤ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

"Where the hell did you get that?" she asked.

"It was my Dad's. I never really knew what to do with it, but tonight seemed like as good a night as any."

She took the flask and sniffed. "Do you even like whiskey?"

"Not really, but it'll be ... an experience! The sting of it sluicing down our throats, it'll

"Do you love me?" she asked. And he answered.

"Well, that is real, isn't it?" she said. And it was.

"We need to leave some mark. Something that will be here when we're not, something that will speak for us."

He rummaged in his pocket and pulled out a penknife. He clicked the large blade into place and walked to the largest fir tree he could see. "What's the plan, Stan?" Polly asked. Howard winked. She watched as he carved his name into the bark. And below it, hers. And between both crooked rows of letters he carved a misshapen, lopsided love-heart.

"There, that's real. That exists and will do when we are gone. I need it to. I need to know that some trace of what we feel will still be here in the morning."

"It will," Polly said. "Of course it will. We feel it don't we? Then it exists. I know you're not religious but, not everything dies when we do." And she said it so that she almost believed it.

"I just want ... I want whatever comes next to know that we were here. And that we felt the way we do. And that it was important to us."

They kissed, and the air above them crackled and ruptured. Prying lights shone through the gaping wormhole in the sky, sickly yellow beams that seemed to come from somewhere beyond the physical world.

They walked to the edge of the cliff hand in hand.

And as the attack ships spilled through into unfamiliar skies, Howard and Polly held each other tight. They glanced back to their names carved on the tree — something real that would remain, they were sure of it. Sure of that much, at least.

They looked into each other's eyes as they let gravity take them over the edge, each holding on to the other, going on their own terms, which was better, and sure that someone, or something, would know how they had felt.

The cities fell.
And the trees burned.
And no one ever knew. ■

Martin Hayes can't write a love story without it turning into a mass-suicide, alien invasion terror-tale. *Opinions as to what this may say about his psychological make-up can be sent to the author at www.paroneiria.com.*



JACEY

be real, won't it?"

Polly flashed a half-smile that made Howard's heart sing. She slugged and grimaced, and handed the flask back, coughing. Howard chuckled and took a deep swig. He felt the world spin a little as it hit the pit of his stomach. He stepped towards the edge and said: "Thanks, Dad. I miss you." Pulling his arm back, he threw the silver flask as high and far as he could. It arced into the murky half-light of the ravine and disappeared.

"Yeah, that was real," Howard said. "But what does that even mean any more? This shouldn't be happening, not in the real world."

Polly shivered and Howard hugged her, he rubbed her arms and did his best to fight back the impotent fury that wanted to erupt and overwhelm him. They both did. They stifled their screams and did not beat their fists against the rocks. For they were better than that, here at the finish, better than the others who were coming to take what was not theirs. They would face the awful end in each other's arms, and on their own terms.

"I just ... it's not real, is it? Is any of this real?"